**PCT** WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68, G01N 33/53 — A1

(11) International Publication Number: **WO 98/23781**

(43) International Publication Date: 4 June 1998 (04.06.98)

(21) International Application Number: PCT/US97/21861

(22) International Filing Date: 26 November 1997 (26.11.97)

(30) Priority Data:

| | | |
|---|---|---|
| 60/031,793 | 26 November 1996 (26.11.96) | US |
| 60/043,560 | 15 April 1997 (15.04.97) | US |

(71) Applicant *(for all designated States except US)*: JOHNS HOPKINS UNIVERSITY [US/US]; 720 Rutland Avenue, Baltimore, MD 21205 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: LI, Min [CN/US]; 8610 N. Field Creek, Lutherville, MD 21093 (US). STRICKER, Nicole, L. [US/US]; 2212 Fox Hunt Lane, Lutherville, MD 21093 (US). BREDT, David, S. [US/US]; 425 Burnett Avenue, San Francisco, CA 94131 (US). CHRISTOPHERSON, Karen, S. [US/US]; 326 Irving Street, San Francisco, CA 94122 (US).

(74) Agents: CORLESS, Peter, F. et al.; Dike, Bronstein, Roberts & Cushman, LLP, 130 Water Street, Boston, MA 02109 (US).

(81) Designated States: AU, CA, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

**Published**
*With international search report.*

(54) Title: LIGAND DETECTION SYSTEM AND METHODS OF USE THEREOF

(57) Abstract

The present invention relates to novel ligand detection systems and methods of using the systems to identify ligands capable of specifically binding orphan protein domains. The invention also relates to peptide ligands capable of specifically binding an orphan domain of interest such as the PDZ domain of neuronal nitric oxide synthase (nNOS). Further provided are methods of detecting the peptide ligands and those orphan protein domains capable of specifically binding the peptide ligands. The present invention is useful for a variety of applications including detecting peptide ligands with therapeutic capacity to treat human diseases.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **AL** | Albania | **ES** | Spain | **LS** | Lesotho | **SI** | Slovenia |
| **AM** | Armenia | **FI** | Finland | **LT** | Lithuania | **SK** | Slovakia |
| **AT** | Austria | **FR** | France | **LU** | Luxembourg | **SN** | Senegal |
| **AU** | Australia | **GA** | Gabon | **LV** | Latvia | **SZ** | Swaziland |
| **AZ** | Azerbaijan | **GB** | United Kingdom | **MC** | Monaco | **TD** | Chad |
| **BA** | Bosnia and Herzegovina | **GE** | Georgia | **MD** | Republic of Moldova | **TG** | Togo |
| **BB** | Barbados | **GH** | Ghana | **MG** | Madagascar | **TJ** | Tajikistan |
| **BE** | Belgium | **GN** | Guinea | **MK** | The former Yugoslav Republic of Macedonia | **TM** | Turkmenistan |
| **BF** | Burkina Faso | **GR** | Greece | | | **TR** | Turkey |
| **BG** | Bulgaria | **HU** | Hungary | **ML** | Mali | **TT** | Trinidad and Tobago |
| **BJ** | Benin | **IE** | Ireland | **MN** | Mongolia | **UA** | Ukraine |
| **BR** | Brazil | **IL** | Israel | **MR** | Mauritania | **UG** | Uganda |
| **BY** | Belarus | **IS** | Iceland | **MW** | Malawi | **US** | United States of America |
| **CA** | Canada | **IT** | Italy | **MX** | Mexico | **UZ** | Uzbekistan |
| **CF** | Central African Republic | **JP** | Japan | **NE** | Niger | **VN** | Viet Nam |
| **CG** | Congo | **KE** | Kenya | **NL** | Netherlands | **YU** | Yugoslavia |
| **CH** | Switzerland | **KG** | Kyrgyzstan | **NO** | Norway | **ZW** | Zimbabwe |
| **CI** | Côte d'Ivoire | **KP** | Democratic People's Republic of Korea | **NZ** | New Zealand | | |
| **CM** | Cameroon | | | **PL** | Poland | | |
| **CN** | China | **KR** | Republic of Korea | **PT** | Portugal | | |
| **CU** | Cuba | **KZ** | Kazakstan | **RO** | Romania | | |
| **CZ** | Czech Republic | **LC** | Saint Lucia | **RU** | Russian Federation | | |
| **DE** | Germany | **LI** | Liechtenstein | **SD** | Sudan | | |
| **DK** | Denmark | **LK** | Sri Lanka | **SE** | Sweden | | |
| **EE** | Estonia | **LR** | Liberia | **SG** | Singapore | | |

# LIGAND DETECTION SYSTEM AND METHODS OF USE THEREOF

## CROSS-REFERENCE TO RELATED APPLICATIONS

The present application is a continuation-in-part of U.S. provisional application serial number 60/031,793, filed November 26, 1996, and U.S. provisional application serial number 60/043,560, filed April 15, 1997, both of which provisional applications are fully incorporated herein by reference.

## BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to a novel ligand detection system and methods of using the system to identify ligands capable of specifically binding orphan protein domains. The invention particularly relates to peptide ligands capable of specifically binding an orphan domain such as the PDZ domain of neuronal nitric oxide synthase (nNOS). Further provided are methods of detecting the peptide ligands and orphan protein domains capable of specifically binding the peptide ligands. The present invention is useful for a variety of applications including detecting peptide ligands with therapeutic capacity to treat human diseases.

Thirteen billion distinct peptides were screened to determine that the nNOS-PDZ domain binds with nanomolar affinity to peptides ending Asp-X-Val. Preference for Asp at the -2 peptide position is mediated by Tyr-77 of nNOS and mutating this residue to His changes the binding specificity from Asp-X-Val to Thr-X-Val. Guided by the Asp-X-Val consensus, candidate nNOS interacting proteins have been identified including glutamate and melatonin receptors. The peptides comprising the consensus sequence Asp-X-Val are useful in altering the interaction of the nNOS PDZ domain with its cognate interacting proteins to prevent the overproduction of NO. Altering the interaction between these proteins with the peptides of the invention can be used to treat many neurodegenerative diseases, including stroke, ALS, Alzheimer's disease, Parkinson's disease and Huntington's disease. The peptides will also be useful for the treatment of muscular dystrophies such as Duchenne muscular dystrophy and motility disorders such as irritable bowel syndrome.

The present invention also relates to a method of identifying the amino acid sequence of a peptide or protein that interacts with a protein domain of interest

(orphan protein domain). The disclosed Protein Interaction Network (PIN) uses an *in vitro* selection strategy that identifies the amino acid sequences which interacts with a given orphan protein domain. This sequence information is then used to search nucleic acid and protein sequence libraries. Interacting PINs from different orphan protein domains are assembled into an electronic resource that can be searched with the sequence of a protein domain of interest.

2. Background

All publications and patent applications herein are incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

A fundamental area of inquiry in biology is the analysis of interactions between proteins. Proteins are complex macromolecules made up of covalently linked chains of amino acids. Each protein assumes a unique three dimensional shape determined principally by its sequence of amino acids. Many proteins consist of smaller units termed domains, which are continuous stretches of amino acids able to fold independently from the rest of the protein. Some of the important functions of proteins are as enzymes, polypeptide hormones, nutrient transporters, structural components of the cell, hemoglobins, antibodies, nucleoproteins, and components of viruses.

Protein-protein interactions enable two or more proteins to associate. A large number of non-covalent bonds form between the proteins when two protein surfaces are precisely matched, and these bonds account for the specificity of recognition. Protein-protein interactions are involved, for example, in the assembly of enzyme subunits; in antigen-antibody reactions; in forming the supramolecular structures of ribosomes, filaments, and viruses; in transport; and in the interaction of receptors on a cell with growth factors and hormones. Products of oncogenes can give rise to neoplastic transformation through protein-protein interactions. For example, some oncogenes encode protein kinases whose enzymatic activity on cellular target proteins leads to the cancerous state. Another example of a protein-protein interaction occurs when a virus infects a cell by recognizing a polypeptide receptor on the surface, and this interaction has been used to design antiviral agents.

Evidence has accumulated over the past years that protein-protein interactions are often mediated by protein modules or domains such as src homology domain 2

(SH2) and src homology domain 3 (SH3). Recently a novel modular domain has been identified in a diverse set of proteins that are typically associated with cell junctions, including synapses of the central nervous system. These novel modular domains are known as PDZ domains. PDZ domains have also been called "GLGF repeats" and Òdisks-large homology repeats" (DHRs) and consist of about 80 amino acids. These domains were first identified as repeated sequences in the neuron-specific postsynaptic density protein (PSD-95/SAP-90), the Drosophila septate junction protein discs-large (dlg), and the epithelial tight-junction protein zona occludens-l (ZOl) (K. Cho et al. *Neuron*, **9**:929-942 (1992); S. Gomperts, *Cell*, **84**:659-662 (1996)). PDZ domains occur in structural proteins of the cytoskeleton and in a heterogeneous family of enzymes that associate with the cytoskeleton, suggesting a role for PDZ domains in protein-protein interactions (C. Ponting et al., *Trends in Biological Sciences*, **20**:102-103 (1995)). Supporting this notion, the three PDZ domains within PSD-95 were first shown to bind the carboxy-terminal Ser/Thr-X-Val motif found in certain N-methyl-D-aspartate (NMDA) type glutamate receptors and in Shaker type potassium channel subunits (E. Kim et al., *Nature*, **378**:85-88 (1995); H. Kornau et al., *Science*, **269**:1737-1740 (1995)). Clustering and localizing channels at synaptic sites is one function of the concatenated domains (M. Sheng, *Neuron*, **17**:575-578 (1996)).

The crystal structures of the third PDZ domains of PSD-95 and dlg have been determined (D. Doyle et al., *Cell*, **85**:1067-1076 (1996); J. Cabral et al., *Nature*, **382**:649-652 (1996)). The PDZ structures show a "carboxylate binding loop", containing the signature GLGF sequence, which interacts with the C-terminal carboxylate group of the peptide ligand. The peptide ligand forms main chain interactions with backbone amide groups in a conserved helix and b strand of the PDZ domain. A critical sequence-specific interaction has been noted between the threonine at the -2 position of the bound peptide and a histidine residue in the PDZ domain (D. Doyle et al., *Cell*, **85**:1067-1076 (1996)). This histidine is conserved in all PDZ repeats of dlg, PSD-95 and related proteins. This histidine, however, is not conserved in other PDZ domains (C. Ponting et al., *Trends in Biological Sciences*, **20**:102-103 (1995)) suggesting distinct peptide-binding specificities.

Since PDZ domains mediate specific protein-protein interactions, critical information in understanding the biological function of PDZ containing proteins is to

determine physiological ligand(s) for orphan PDZ domains. Recent evidence shows that interaction between the PDZ domain and peptide ligands can be regulated by differential affinity (B. Muller et al., *Neuron*, **17**:255-265 (1996)) and by protein phosphorylation (N. Cohen *Neuron*, **17**:759-767 (1996)). These mechanisms, however, are not adequate to explain the diversity of PDZ-target protein interactions in both excitable and non-excitable tissues.

Nitric oxide (NO), an endogenous signaling molecule, plays critical roles in nervous, immune, and cardiovascular function (D. Bredt et al., *Ann. Rev. Biochem.*, **63**:175-195 (1994); M. Marletta, *J. Biol. Chem.*, **268**:12231-12234 (1993); S. Moncada et al., *N. Eng. J. Med.*, **329**:2002-2012 (1993)). Physiological studies have demonstrated numerous functions for neuron-derived NO, produced primarily by the neuronal NO synthase (nNOS). However, excess nNOS activity mediates brain injury in cerebral ischemia and in animal models of Parkinson's disease (T. Dawson et al., *Ann. Neurol.*, **32**:297-311 (1992); P. Hantraye et al., *Nature Medicine*, **2**:1017-1021 (1996); Z. Huang et al., *Science*, **265**:1883-1885 (1994)). Excess nNOS activity has been broadly linked with many neurodegenerative diseases, motility disorders and muscular dystrophies, including Alzheimer's disease, Huntington's disease (see generally D. Bredt et al., *Nature*, **351**:714-718 (1991)). nNOS activity must therefore be tightly regulated. One level of regulation is reflected by molecular targeting of the nNOS to specific intracellular membrane domains (C. Aoki et al., *Brain Res.*, **620**:97-113 (1993)). This subcellular localization is mediated by the N-terminus of nNOS, which contains a PDZ domain (J. Brenman et al., *Cell*, **82**:743-752 (1995)). This N-terminal domain of nNOS interacts with the PDZ domain of a 1-syntrophin and the second PDZ domains of PSD-95 and PSD-93. These interactions target nNOS to synaptic sites in skeletal muscle and brain (J. Brenman et al., *Cell*, **84**:757-767 (1996)). The structural details of these PDZ-PDZ interactions are not yet known.

Several lines of evidence suggest that additional binding partners for the PDZ domain of nNOS may also exist. First, not all membrane-associated nNOS in brain is bound to PSD-95 and related proteins (J. Brenman et al., *Journal of Neuroscience*, (1996) (in press) unpublished observations). Also, in certain muscle diseases, nNOS does not interact properly with a 1-syntrophin at the skeletal muscle sarcolemma (D. Chao et al., *Journal of Experimental Medicine*, **184**:609-618 (1996)). We therefore sought to determine whether specific carboxylate-peptides might also associate with

the PDZ domain of nNOS. Identification of such peptides would facilitate the structure and function study of PDZ domains. Also, the *in vitro* defined peptide sequences may help identify additional nNOS interacting proteins.

Protein-protein interactions have been generally studied in the past using biochemical techniques such as cross-linking, co-immunoprecipitation and co-fractionation by chromatography. One of the disadvantages of these techniques is that interacting proteins often exist in very low abundance and are, therefore, difficult to detect. Another major disadvantage is that these biochemical techniques involve only the proteins, not the genes encoding them. When an interaction is detected using biochemical methods, the newly identified protein often must be painstakingly isolated and then sequenced to enable the gene encoding it to be obtained. Another disadvantage is that these methods do not immediately provide information about which domains of the interacting proteins are involved in the interaction.

*In vitro* determination of ligands for peptide-binding domains, such as 5H3 and SH2 motifs, has been achieved using two types of random peptide libraries (A. Sparks et al., *Methods Enzymol.*, **255**:498-509 (1995); M. Sheng, *Neuron*, **17**:575-578 (1996); S. Zhou et al., *Methods Enzymol.*, **254**:523-535 (1995); and review by M. Gallop et al., *Journal of Medicinal Chemistry*, **37**:1233-1251 (1994)). One strategy utilizes the filamentous phage coat protein to display random N-terminal peptides. By repeated rounds of affinity panning and amplification, individual interacting peptides can be identified by sequencing the corresponding coding region of phage DNA (A. Sparks et al., *Methods Enzymol.*, **255**:498-509 (1995)). A second approach uses soluble random peptides that are chemically synthesized. By affinity purification of a mixture of bound peptides and subsequent peptide sequencing, a population based consensus can be deduced (S. Zhou et al., *Methods Enzymol.*, **254**:523-535 (1995)). Because the phage display system only accommodates N-terminal peptides, it can not be used to select C-terminal peptides for the PDZ domain. Although chemical peptide libraries are applicable, the approach has difficulties in accommodating cysteine and tryptophan and does not provide individual ligand sequences. As a result, analyses of chemical libraries cannot resolve compensatory effects potentially present in peptides of low abundance and may miss high affinity sequences containing tryptophan and cysteine. Thus, it would be desirable to use a genetic strategy to screen a large pool

of C-terminal peptides containing all 20 amino acids to identify individual PDZ binding peptides.

SUMMARY OF THE INVENTION

In one aspect, the invention relates to peptides capable of altering the interaction between the nNOS PDZ domain and the proteins which this domain interacts. The peptides preferably alter the interactions between the nNOS PDZ domain and melatonin or non-NMDA type glutamate receptors. The peptides of the invention are useful in the formulation of therapeutic compositions which alter intermolecular binding between the nNOS PDZ domain and the proteins which this domain interacts *in vivo*. Via inhibition -of these interactions, the peptides of the invention will be useful in suppressing the production of excess levels of NO which are neurotoxic and contribute to myofiber necrosis. For example, the peptides of the invention can be used to treat many neurodegenerative diseases, including stroke, ALS, Alzheimer's disease, Parkinson's disease and Huntington's disease. The peptides are also useful for the treatment of muscular dystrophies such as Duchenne muscular dystrophy and motility disorders such as irritable bowel syndrome.

Another aspect of the invention is to provide peptides comprising the general sequence D-X-V-COOH wherein D=Aspartic acid, X is variable and V=Valine.

Another object of the invention is to provide peptides capable of altering the interaction between the nNOS PDZ domain and the proteins which this domain interacts which are useful as commercial laboratory or bioprocess reagents.

Another object of the invention is to provide peptides which can be used as molecular probes that specifically label nNOS. For instance, the peptides of the invention can be labeled according to standard procedures in the art and can be used as molecular probes to detect nNOS *in vivo* or *in vitro*.

The invention also provides a kit comprising peptides which interact with the PDZ domain of nNOS.

Another aspect of the invention is isolated nucleic acid sequences that encode the peptides described herein.

Another object of the invention is to couple a genetic system that identifies peptides which interact with a given protein domain (orphan protein domain) with the available electronic sequence databases. The genetic system provides the sequence of the peptide which interacts with the orphan protein domain. This sequence is then

used to identify proteins already present in the electronic nucleic acid and protein sequence databases. A Protein Interaction Network (PIN) is then assembled which correlates the peptide sequences which interact with a given orphan protein domain. Assembly of many different PINs results in the assembly of a Super Protein Interaction Network (SPIN) which will serve as an electronic extension for existing sequence databases. This allows the researcher to search the database with the sequence of a given orphan protein domain for peptide sequences which are known to specifically interact with a given orphan protein domain.

The invention also relates to a peptide ligand detection system that includes a random peptide library preferably of at least about $10^6$ members comprising a recombinant DNA vector encoding a DNA binding protein. The DNA binding protein is selected to specifically bind a DNA sequence on the vector. The DNA binding protein encoded by the DNA vector comprises a random peptide sequence covalently linked to the DNA binding protein as an in-frame fusion protein. The fusion protein is typically formatted so that the DNA vector can encode preferably at least about $10^6$ different fusion proteins up to about $10^8$ fusion proteins or more, each of which is capable of specifically binding the DNA sequence on the vector. The peptide ligand detection system further includes an orphan protein domain sequence immobilized on a solid support that is capable of specifically binding the random peptide of the DNA binding protein.

Significantly, the ligand detection system of the present invention utilizes an immobilized orphan protein domain sequence to specifically bind the random peptide of the in-frame fusion protein. Typically, the orphan protein domain sequence is a contiguous or non-contiguous amino acid sequence within the linear sequence of a protein of interest. Sometimes the orphan protein domain sequence is referred to as a protein module. In contrast, prior ligand detection systems using random peptide libraries rely on substantially larger molecules to bind the ligand, e.g., receptors, antibodies, or enzymes. Exemplary orphan protein domain sequences are illustrated below in Figure 7.

The peptide ligand detection system can further include an inducer molecule capable of specifically binding the DNA binding protein. Typically, the inducer molecule is selected to release the recombinant DNA vector from the immobilized

orphan protein domain sequence. In particular, the inducer molecule can be isopropylthio-β-D-galactoside (IPTG).

A peptide ligand detection system in accord with the present invention can include one of a variety of suitable recombinant DNA vectors. That is, the recombinant DNA vectors can encode a variety of suitable DNA binding proteins and DNA sequences capable of being bound by the DNA binding proteins.

For example, the DNA binding protein of the peptide ligand detection system can include a prokaryotic repressor protein sequence. In addition, the DNA sequence bound by the DNA binding protein can be a prokaryotic operator sequence. More specifically, the prokaryotic repressor protein sequence can be a lac repressor or a fragment thereof capable of specifically binding the DNA sequence on the recombinant DNA vector. In addition, the prokaryotic operator sequence can be lac O or a fragment thereof capable of being specifically bound by the prokaryotic repressor protein sequence.

As noted, the recombinant DNA vectors of the random peptide library are formatted to express the random peptide as a fusion protein. A DNA binding protein of the invention typically features high avidity binding to DNA and has a region preferably at the C-terminus of the protein that can accept an amino acid sequence insertion without interfering with the DNA binding activity of the protein. The half-life of a specific binding pair formed between the DNA binding protein and the recombinant DNA vector must be long enough for screening to occur. In general, that half-life will be at least about one to four hours or longer. The half-life of the specific binding pair formed between the random peptide and the immobilized orphan protein domain will also be about one to four hours or longer.

If desired, the peptide ligand detection system can include an in-frame peptide linker sequence, e.g., between the prokaryotic repressor protein sequence (or fragment) and the random peptide sequence.

A peptide ligand detected by the present ligand detection system is capable of specifically binding the immobilized orphan protein domain of interest. The binding affinity ($EC_{50}$) of the specific binding interaction depends on several parameters such as the degree of binding affinity desired and the complexity of the random peptide sequence. However, in general the binding affinity will be in the micromolar to nanomolar range for most immobilized orphan protein domains.

As will be discussed more fully below, an exemplary peptide ligand in accord with the present invention comprises between about 3, 6, 7, 8, 9, 10, 12, 15, 20, 25, 30, 35, 40, 50 or more amino acids. For example, the present invention provides a peptide ligand comprising about 3, 6, 7, 8, 9 or 10 amino acids in which the C-terminal sequence of the peptide ligand consists of the sequence D-X-V-COOH, wherein D is Asp, X is any amino acid, preferably any of the 20 common natural amino acids, and V is Val. That peptide ligand has been found to specifically bind a specified orphan protein domain (PDZ).

In general, a peptide ligand in accord with the invention has a binding affinity ($EC_{50}$) for an orphan protein domain preferably in the micromolar to nanomolar range. Preferred peptide ligands have an $EC_{50}$ in the nanomolar range.

In particular, the immobilized orphan protein domain can be a PDZ domain such as those obtained from a variety of known proteins such as nitric oxide synthase (nNOS), post-synaptic density protein (PSD-95/SAP-90), post-synaptic density protein (PSD-93), epithelial tight-junction protein zona occludens-1 (ZO1), N-methyl-D-aspartate (NMDA) type glutamate receptor, Shaker-type potassium channel subunit, and 1-syntrophin.

The invention further provides therapeutic compositions comprising a peptide ligand of the present invention. The therapeutic compositions are preferably provided in a pharmaceutically acceptable vehicle, e.g. sterile and pyrogen-free. Examples of preferred therapeutic compositions are specified below.

Further provided are isolated nucleic acids encoding peptide ligands of the present invention and particularly DNA vectors comprising the isolated nucleic acids.

The present invention also provides a method of detecting a peptide ligand capable of specifically binding an orphan protein domain of interest. In general, the method includes lysing transformed cells comprising the random peptide library generally discussed above. The lysing is under conditions such that the DNA binding protein comprising the random peptide remains bound to the recombinant DNA vector. The method further includes the steps of contacting the fusion proteins of the random peptide library to an immobilized orphan protein domain under conditions conducive to specific peptide-orphan protein domain binding and isolating a recombinant DNA vector encoding a fusion protein that specifically binds to the orphan protein domain.

In most cases, the method will further include the steps of transforming a host cell with the isolated recombinant DNA vector obtained, repeating the lysing and contacting steps and isolated a selected recombinant DNA vector. As will be shown below in the examples, practice of this method leads to amplification of the selected recombinant DNA vector.

The method will also typically includes the steps of determining the amino acid sequence of the random peptide encoded by the selected recombinant DNA vector, and searching a protein sequence database to identify an orphan protein domain in the database comprising the random peptide.

If desired, the method can further include the step of assembling a protein interaction network (PIN) sufficient to correlate (particularly match) a plurality of random peptide sequences to the orphan protein domain. In this method, the plurality of random peptide sequences are capable of binding the correlated orphan protein domain with a binding affinity in the micromolar to nanomolar range as noted below.

The method can further include assembling a super protein interaction network (SPINS) comprising a plurality of protein interaction networks (PINs) sufficient to serve as an electronic extension database for the protein sequence database.

Typically, the assembly is assisted by one or more suitable computer programs such as those generally known in the field for compiling protein and/or nucleic sequences in a matrix or matrix-type format. The matrix or matrix-type format can be readily searched with a test sequence that can be, e.g., a peptide ligand sequence or orphan domain sequence in accord with the invention.

The invention further provides a method of detecting a peptide ligand capable of specifically binding an orphan protein domain of interest, the method comprising searching the super protein interaction network (SPINS) with an amino acid sequence comprising an orphan protein domain of interest, and identifying the peptide ligand capable of specifically binding the orphan protein domain of interest. The peptide ligand can be obtained from any suitable source such as any of the random peptide libraries discussed previously.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic diagram showing affinity selection from a C-terminal peptide library.

Figures 2A is a graph showing affinity selection of peptides interacting with PDZ3 of PSD-95 by ELISA.

Figure 2B is an alignment of deduced amino acid sequences of PDZ3 specific clones. Eleven clones were randomly chosen and sequenced. Single letter code for 20 amino acids are used. Italic letters indicate amino acids present at the end of the linker which separates Lac I and the fused peptide. "*" indicates a stop codon.

Figure 3A is a graph showing *in vitro* selection of peptides interacting with nNOS-PDZ. The graph shows identification of nNOS-PDZ interacting clones by ELISA. After 4 rounds of affinity panning, a total of 150 individual clones were randomly selected and tested for interaction with nNOS-PDZ by ELISA as described in experimental procedures. Clones 1 to 48 are shown (horizontal axis). Gray bars: BSA; open bars: GST-$NAB_{HERG}$ + BSA; closed bars: GST-nNOS-PDZ+ BSA.

Figure 3B and 3C illustrate a sequence alignment of 95 independent nNOS binding peptides (NBPs). The deduced amino acid sequence of the clones were obtained and aligned according to the first stop codon (*). The italic Gs are part of linker region. The library template (*GGG*-$X_{15}$-*) is shown at the top of the sequence alignment.

Figures 4A -4I are graphs showing determinations of a consensus nNOS binding peptide (NBP). Normalized amino acid abundance of the final nine residues from the population of 95 independent nNOS binding peptides (closed bars) is compared in each figure with codon frequency in the original library (open bars). Residues in the library linker region were not included in each figure.

Figure 5A is a graph showing all 95 NBPs fail to interact with PDZ3. ELISA results of 36 randomly chosen NBP clones are shown. Horizontal axis: NBP clone number; vertical axis: ELISA signal normalized against clones with strongest binding.

Figure 5B is a graph illustrating that mutating Y77D78 to H77E78 changes the nNOS PDZ binding specificity from D-X-V to T-X-V. ELISA results of two high affinity peptides are shown. NBP-161 for nNOS ($EC_{50}$=~8 nM) and PD-325 for PDZ3 (EC=~2 nM) are expressed as maltose binding protein fusion and affinity purified on amylose agarose beads (see Experimental Procedures).

Figure 5C is a graph showing that the aspartate at the -2 position is critical for NBP binding. Single amino acid substitutions at the -2 position were obtained. The

peptides were expressed at maltose binding protein fusion at the C-terminus (see Experimental procedures). ELISA results of seven mutants are shown.

Figure 5D is a representation of a Western immunoblot. Solubilized brain extracts were incubated with amylose resin alone (lane 1), amylose resin saturated with a maltose binding protein fusion containing a C-terminal NPB-123 (lane 2) or with the same fusion protein in which the -2 aspartate was changed to threonine (lane 3). The beads were washed and retention of nNOS was detected by western blotting. Molecular weight standards in kDa are marked on the left.

Figure 6 is a schematic diagram showing that functional nNOS PDZ has a uniquely large structure. The location of the PDZ domain is shown in the N-terminus of nNOS. Interaction of nNOS with the PDZ domains of PSD-93 requires amino acids 16-130 of nNOS. Association of nNOS fusions with PSD-93 was evaluated by the yeast two hybrid system and is expressed as β-galactosidase units. Interactions of five different NBPs (#64-68) with nNOS fusions were evaluated by ELISA and is expressed as normalized 0D405.

Figure 7 is a list of known orphan protein domains (common protein modules).

Figures 8A-8R show results of search (scan) of a non-redundant protein sequence database (Genbank) identifying protein sequences comprising the -D-X-V-COOH sequence where D is Asp, X is any of the 20 common amino acids, and V is Val. Identified protein sequences are listed in bold script and are grouped according to species (human, mouse, rat, etc.). Various descriptors accompany each identified protein sequence in accord with nomenclature adopted by Genbank.

DETAILED DESCRIPTION OF THE INVENTION

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. For purposes of the present invention, the following terms are defined below.

In the polypeptide notation used herein, the left-hand direction is the amino terminal direction and the right-hand direction is the carboxy-terminal direction, in accordance with standard usage and convention. Similarly, unless specified

otherwise, the left-hand end of single-stranded polynucleotide sequences is-the 5' end; the left-hand direction of double-stranded polynucleotide sequences is referred to as the 5' direction. The direction of 5' to 3' addition of nascent RNA transcripts is referred to as the transcription direction; sequence regions on the DNA strand having the same sequence as the RNA and which are 5' to the 5' end of the RNA transcript are referred to as "upstream sequences"; sequence regions on the DNA strand having the same sequence as the RNA and which are 3' to the 3' end of the RNA transcript are referred to as "downstream sequences".

The term "protein interaction inhibitor" is used herein to refer to an agent which is identified by one or more screening method(s) of the invention as an agent which selectively inhibits protein-protein binding between a first interacting polypeptide and a second interacting polypeptide. Some protein interaction inhibitors may have therapeutic potential as drugs for human use and/or may serve as commercial reagents for laboratory research or bioprocess control. Protein interaction inhibitors which are candidate drugs are then tested further for activity in assays which are routinely used to predict suitability for use as human and veterinary drugs, including *in vivo* administration to non-human animals and often including administration to human in approved clinical trials.

As used herein, the term "operably linked" refers to a linkage of polynucleotide elements in a functional relationship. A nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For instance, a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the coding sequence.

Operably linked means that the DNA sequences being linked are typically contiguous and, where necessary to join two protein coding regions, contiguous and in reading frame. However, since enhancers generally function when separated from the promoter by several kilobases and intronic sequences may be of variable lengths, some polynucleotide elements may be operably linked but not contiguous.

As used herein, the term "orphan protein domain" refers to any domain of a protein which binds or interacts with another protein, particularly but not limited to PDZ domains. Orphan protein domains are typically contiguous stretches of amino acids that facilitate protein-protein interactions. Orphan protein domains, however, do include domains comprising non-contiguous stretches of amino acids that through

secondary and tertiary structure are brought into association to facilitate protein-protein interactions. Protein-protein interactions typically comprise but are not limited to, non-covalent bonds that account for the specificity of interaction between two proteins. Examples of such non-covalent bonds include van der Waals contacts, hydrogen bonds and salt bridges. Examples of known orphan protein domains are set forth in Figure 7.

Preferred orphan protein domains have a length of between about 1 to 1000 amino acids, preferably about 1 to 500 amino acids, and more preferably about 1 to 100 amino acids. Particularly preferred orphan protein domains include more than one amino acid and are capable of specifically binding a peptide ligand with a binding affinity ($EC_{50}$) of between about 0.001 to 100 μM, preferably 0.2 to 1μM and more preferably 8 to 100 nM as defined by any suitable immunological assay such as Western blotting, ELISA, RIA, gel mobility shift assay, enzyme immunoassay, competitive assays, saturation assays or other suitable protein binding assays known in the field and specified below. See generally Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York (1989), Sambrook et al. *infra*, and Harlow and Lane Antibodies: A Laboratory Manual, CSH Publications, N.Y. (1988), for disclosure relating to suitable methods for detecting specific binding between proteins.

A "DNA binding protein" as the term is used herein, refers to a protein that specifically binds a DNA strand and preferably two DNA strands of the recombinant DNA vector. More preferably, the DNA binding protein specifically binds to the specific DNA sequence included in the vector. In embodiments of the invention in which RNA vectors are used, DNA binding protein can also refer to an RNA binding protein.

Suitable DNA binding proteins are known in the field. For example, suitable prokaryotic DNA binding proteins include lac repressor, phage 434 repressor, lambda phage cI and cro repressors, phage P22 Arc and Mnt repressors, and CAP protein. Also included are eukaryotic DNA binding proteins such as those comprising homoeoboxes with helix-turn-helix motifs, proteins including helix-loop-helix structures particularly myc; fos, jun and other proteins including leucine zippers and DNA binding domains, POU domain proteins, TFIIIA, and yeast Gal4 protein.

Preferably, the DNA binding protein is the lac repressor particularly the 37 kDa protein encoded by E. coli lac I gene capable of repressing transcription from the lacZYA operon by binding to a specific DNA sequence termed lacO. See e.g., Aububel et al. *supra*; Sambrook et al., *supra*; Knight et al. *J. Biol. Chem.* **264**:3639-3642 (1989); Beyreuther in *The Operon* (Miller and Reznikoff, eds. Cold Spring Harbor Laboratory (1980)).

A "host cell" as the term is used herein is a eukaryotic or prokaryotic cell or cell group that is capable of being transformed by a recombinant DNA vector. Preferably, the host cell is a suitable bacterial strain such as E. coli K12.

A "peptide ligand" refers to a molecule and particularly a peptide such as a random peptide that is capable of being specifically bound by an immobilized orphan protein domain. In addition, the peptide ligand is capable of being bound by the orphan protein domain as it exists in a protein. Preferably, the binding affinity ($EC_{50}$) between the peptide ligand and the immobilized orphan protein domain is between about 0.001 to 100 μM, preferably 0.2 to 1μM and more preferably 8 to 100 nM as determined by a suitable binding assay as described herein.

By the term "specific binding" or similar term is meant a molecule disclosed herein which binds another molecule, thereby forming a specific binding pair, but which does not recognize and bind to other molecules as determined by, e.g., Western blotting, ELISA, RIA, gel mobility shift assay, enzyme immunoassay, competitive assays, saturation assays or other suitable protein binding assays known in the field..

By the term "immobilized orphan protein domain" is meant an amino acid sequence corresponding to a desired orphan protein domain that has been covalently or non-covalently bound to a solid support or surface such as a particle or a dish. If desired the immobilized orphan protein can be immobilized by attaching an immunologically recognizable ligand, e.g., biotin, bound to streptavidin which is attached to the solid support or surface. If desired, the ligand may be attached by a peptide linker sequence.

Exemplary peptide linker sequences in accord with the invention comprise up to 20 amino acids, preferably up to about 10 amino acids, and more preferably from about 1 to 5 amino acids. The linker sequence is generally flexible so as not hold the random peptide in a single rigid conformation. The linker sequence can be used, e.g., to space the DNA binding protein from the fused random peptide sequence.

Preferably, the orphan protein domain will be between about 1000, preferably 500 and more preferably 100 amino acids in length. It is also preferred that the orphan protein domain be immobilized on a solid support or surface which is conducive to standard affinity panning (i.e. biopanning or panning) techniques capable of detecting nanomolar binding affinities between proteins. A preferred solid support is a microtitre dish.

The term "random peptide" refers to an amino acid oligomer comprising two or more amino acid residues that have been constructed by a recognized stochastic or random process. A "random peptide library" refers not only to a set of recombinant DNA vectors that encodes a set of random peptides, but also to the set of random encoded by those vectors, as well as the fusion proteins containing those random peptides.

The Protein Interaction Network (PIN) is generally applicable to identifying the amino acid sequences which interact with a given orphan protein domain.

A PIN in accord with the invention can be assembled and then stored in a variety of ways. For example, a desired PIN can be assembled and stored by use of a computer program such as Netscape and particularly a Netscape assisted program. The program can be run (i.e. performed) on any suitable computer such as an PC (IBM) or Macintosh (Apple) computer. A preferred PIN includes between about 100 to $10^{13}$, preferably about 1000 to $10^{12}$, and more preferably about $10^{12}$ peptide ligand sequences.

Once assembled, the PIN of interest can be further assembled into a Super Protein Interaction Network (SPIN) by use of a computer program such as BLAST run on, e.g., a conventional central server system. The size of the SPIN will depend on several parameters such as the complexity of the PIN assembly and desired electronic connections with other database networks. In general, the SPIN will include between about 5 to $10^8$, preferably 500 to $10^8$, and more preferably 500 to $10^7$ PINs. Compilation and analysis of multiple PINs is facilitated by any number of stand alone computer-assisted programs particularly BLAST and other secondary sequence computer programs known in the field.

The present invention is based on the discovery that a random fusion protein library wherein random peptides are fused to the C-terminus of a bacterial DNA binding protein such as a transcriptional repressor can be used to select for specific

peptide ligands that bind to a given orphan domain. The gene encoding the fusion protein is operably linked on a plasmid to the fusion protein's binding site. Following expression or induction of the election of the fusion protein in a transformed or transfected host cell, the fusion protein binds to its cognate binding sequence on the plasmid. This linkage of the fusion protein to the plasmid which itself encodes the fusion protein allows for repeated rounds of selection for specific peptide ligands in the library by affinity purification of fusion protein-plasmid complexes using an orphan domain of interest. The plasmid can then be dissociated from the complex and used to retransform appropriate host cells for another round of selection.

Generally, the nomenclature used hereafter and the laboratory procedures in cell culture, molecular genetics1 and nucleic acid chemistry and cell culture described below are those well known and commonly employed in the art. Standard techniques are used for recombinant nucleic acid methods, polynucleotide synthesis, and microbial culture and transformation (e.g., electroporation, lipofection). Generally enzymatic reactions and purification steps are performed according to the manufacturer's specifications. The techniques and procedures are generally performed according to conventional methods in the art and various general references (see, generally, Sambrook et al., Molecular Cloning: A Laboratory Manual, 2d ed. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., which is incorporated herein by reference) which are provided throughout this document. The procedures therein are believed to be well known in the art and are provided for the convenience of the reader. All the information contained therein is incorporated herein by reference.

General methods for assembling amino acid and nucleic acid sequence data in accord with the methods described herein have been disclosed. See S. Altschul et al. *J. Mol. Biol.*, **215**:403-410 (1990); and S. Altschul et al. *Nuc. Acids Res.*, **25**:3389-3402 (1997) for disclosure relating to the BLAST, particularly gapped BLAST, and PSI-BLAST computer programs the disclosures of which are fully incorporated herein by reference.

Peptides of the invention comprising those that bind to nNOS are at least 3 amino acids long and comprise the consensus sequence Asp-X-Val. Peptides of longer length are also encompassed within the invention with the proviso that the peptide contain the consensus sequence, preferably at the C-terminal end.

Accordingly, peptides of at least 5 amino acids, at least 7 amino acids, at least 10 amino acids and at least 15 or more amino acids are encompassed.

The peptides of the invention may be prepared by recombinant nucleotide expression techniques or by chemical synthesis using standard peptide synthesis techniques. For example, peptides of the invention can be produced, for example, by expressing cloned nucleotide sequences. Alternatively, peptides of the invention can be generated directly from intact protein products. Peptides can be specifically cleaved by proteolytic enzymes, including, but not limited to, trypsin, chymotrypsin or pepsin. Each of these enzymes is specific for the type of peptide bond it attacks. Trypsin catalyzes the hydrolysis of peptide bonds whose carbonyl group is from a basic amino acid, usually arginine or lysine. Pepsin and chymotrypsin catalyze the hydrolysis of peptide bonds from aromatic amino acids, particularly tryptophan, tyrosine and phenylalanine. Alternate sets of cleaved peptide fragments are generated by preventing cleavage at a site which is susceptible to a proteolytic enzyme. For example, reaction of the epsilon -amino groups of lysine with ethyltrifluorothioacetate in mildly basic solution yields a blocked amino acid residue whose adjacent peptide bond is no longer susceptible to hydrolysis by trypsin (Goldberger et al., *Biochem.*, **1**:401 (1962)).

Peptides of the invention also can be modified to create peptide linkages that are susceptible to proteolytic enzyme catalyzed hydrolysis. For example, alkylation of cysteine residues with beta -halo ethylamines yields peptide linkages that are hydrolyzed by trypsin (Lindley, *Nature*, **178**:647 (1956)). In addition, chemical reagents that cleave peptide chains at specific residues can be used (Withcop, *Adv. Protein Chem.*, **16**:221 (1961)). For example, cyanogen bromide cleaves peptides at methionine residues (Gross et al., *J. Am Chem Soc.*, **83**:1510 (1961)). Thus, by treating full-length proteins with various combinations of modifiers, proteolytic enzymes and/or chemical reagents, numerous discrete overlapping peptides of varying sizes are generated. These peptide fragments can be isolated and purified from such digests by chromatographic methods.

Most preferably, isolated peptides of the present invention can be synthesized using an appropriate solid state synthetic procedure (Steward and Young, Solid Phase Peptide Synthesis, Freemantle, San Francisco, Calif. (1968)). A preferred method is the Merrifield process (Merrifield, *Recent Progress in Hormone Res.*, **23**:451 (1967)).

The binding activity of these peptides may conveniently be tested using, for example, the assays as described herein.

Once an isolated peptide of the invention is obtained, it may be purified by standard methods including chromatography (e.g., ion exchange, affinity, and sizing column chromatography), centrifugation, differential solubility, or by any other standard technique for protein purification. For immunoaffinity chromatography, a peptide may be isolated by binding it to an affinity column comprising antibodies that were raised against that peptide, or a related peptide of the invention, and were affixed to a stationary support. Alternatively, affinity tags such as hexa-His (Invitrogen), Maltose binding domain (New England Biolabs, Inc.), influenza coat sequence (Kolodziej et al., *Methods Enzymol.*, **194**:508-509 (1991)), and glutathione-S-transferase can be attached to the peptides of the invention to allow easy purification by passage over an appropriate affinity column. A DNA affinity column using DNA containing a sequence encoding the peptides of the invention could be used in purification.

Isolated peptides can also be physically characterized using such techniques as proteolysis, nuclear magnetic resonance, and x-ray crystallography.

With regard to nucleic acid sequences of the present invention, "isolated" means: an RNA or DNA polymer, portion of genomic nucleic acid, cDNA, or synthetic nucleic acid which, by virtue of its origin or manipulation:

(i) is not associated with all of a nucleic acid with which it is associated in nature (e.g. is present in a host cell as a portion of an expression vector); or

(ii) is linked to a nucleic acid or other chemical moiety other than that to which it is linked in nature; or

(iii) does not occur in nature.

By "isolated" it is further meant a nucleic acid sequence:

(i) amplified *in vitro* by, for example, polymerase chain reaction (PCR);

(ii) synthesized by, for example, chemical synthesis;

(iii) recombinantly produced by cloning; or

(iv) purified, as by cleavage and gel separation.

The nucleic acid sequences of the present invention may be characterized, isolated, synthesized and purified using no more than ordinary skill. See Sambrook et

al., Molecular Cloning, Cold Spring Harbor Press, New York, 1989,incorporated herein by reference.

Due to the degeneracy of nucleotide coding sequences (see Alberts et al., Molecular Biology of the Cell, Garland Publishing, New York and London, 1989-page 103, incorporated herein by reference), a number of different nucleic acid sequences may be used in the practice of the present invention. These include, but are not limited to, sequences encoding the peptides of Figure 3B and 3C. This includes the substitution of different codons encoding the same amino acid residue within the sequence, thus producing a silent change. Almost every amino acid except tryptophan and methionine is represented by several codons. Often the base in the third position of a codon is not significant, because those amino acids having 4 different codons differ only in the third base. This feature, together with a tendency for similar amino acids to be represented by related codons, increases the probability that a single, random base change will result in no amino acid substitution or in one involving an amino acid of similar character.

The nucleotide sequences of the invention can be altered by mutations such as substitutions, additions or deletions that provide for functionally equivalent nucleic acid sequence. In particular, a given nucleotide sequence can be mutated *in vitro* or *in vivo*, to create variations in coding regions and/or to form new restriction endonuclease sites or destroy preexisting ones and thereby to facilitate further *in vitro* modification. Any technique for mutagenesis known in the art can be used including, but not limited to, *in vitro* site-directed mutagenesis (Hutchinson et al., *J. Biol. Chem.*, **253**:6551 (1978)), use of TAB Registered TM linkers (Pharmacia), PCR-directed mutagenesis, and the like. The functional equivalence of such mutagenized sequences, as compared with unmutagenized sequences, can be empirically determined by comparisons of structural and/or functional characteristics.

The isolated nucleotide sequences of the invention may be cloned or subcloned using any method known in the art (See, for example, Sambrook, J. et al., Molecular Cloning, Cold Spring Harbor Press, New York, 1989), the entire contents of which are incorporated herein by reference. In particular, nucleotide sequences of the invention may be cloned into any of a large variety of vectors. Possible vectors include, but are not limited to, cosmids, plasmids or modified viruses, although the vector system must be compatible with the host cell used. Viral vectors include, but

are not limited to, lambda, simian virus, bovine papillomavirus, Epstein-Barr virus, and vaccinia virus. Viral vectors also include retroviral vectors, such as Amphatrophic Murine Retrovirus (see Miller et al., *Biotechniques*, 7:980-990 (1984)), incorporated herein by reference). Plasmids include, but are not limited to, pBR, PUC, pGEM (Promega), and Bluescript Registered TM (Stratagene) plasmid derivatives. Introduction into and expression in host cells is done for example by, transformation, transfection, infection, electroporation, etc.

Examples of DNA vectors for constructing random peptide libraries, methods of making same, and useful related materials and methods have been disclosed in U.S. Pat. Nos. 5,270,170 and 5.498,530, the disclosures of which are incorporated herein by reference.

The peptides described herein can be used in pharmaceutical compositions to alter the binding of the nNOS PDZ domain and the proteins which this domain interacts. The peptides preferably alter the interactions between the nNOS PDZ domain and melatonin or non-NMDA type glutamate receptors. An exemplary pharmaceutical composition is a therapeutically effective amount of one of the disclosed peptides optionally included in a pharmaceutically-acceptable and compatible carrier. The term "pharmaceutically-acceptable and compatible carrier" as used herein, and described more fully below, refers to one or more compatible solid or liquid filler diluents or encapsulating substances that are suitable for administration to a human or other animal. In the present invention, the term "carrier" thus denotes an organic or inorganic ingredient, natural or synthetic, with which the peptides of the invention are combined to facilitate administration.

Peptides of the invention can be stabilized to decrease protease sensitivity and/or increase *in vivo* half-life by methods known in the art. For instance, peptides of the invention can be modified by the addition of a N or C terminal tail, modified by the methylation or glyoxylation of the termini or by substitution or other modification to the sequence to increase the peptide half-life, stability, and/or protease resistance.

In some embodiments, the peptides are conformationally restricted such as those which are cyclicized, circularized or otherwise restricted by peptide and/or non-peptide bonds to limit conformational variation and/or to increase stability and/or half-life of the peptides. In some embodiments, peptides are provided as linear peptides.

In some embodiments, peptides of the present invention comprise one or more D amino acids. As used herein, the term "D amino acid peptides" is meant to refer to peptides according to the present invention which comprise at least one and preferably a plurality of D amino acids. D amino acid peptides consist of 4-25 amino acids. D amino acid peptides retain the biological activity of the peptides of the invention that consist of L amino acids, i.e. D amino acid peptides inhibit the interaction of nNOS and the proteins which bind to nNOS. In some embodiments, the use of D amino acid peptides is desirable as they are less vulnerable to degradation and therefore have a longer half life. D amino acid peptides comprising mostly all D amino acids or D amino acid peptides that consist of only D amino acids may comprise amino acid sequences in the reverse order of amino acid sequences of peptides.

The term "therapeutically-effective amount" is that amount of the present pharmaceutical compositions which produces a desired result or exerts a desired influence on the particular condition being treated. Various concentrations may be used in preparing compositions incorporating the same ingredient to provide for variations in the age of the patient to be treated, the severity of the condition, the duration of the treatment and the mode of administration.

The term "compatible" as used herein, means that the components of the pharmaceutical compositions are capable of being commingled with the peptides of the present invention, and with each other, in a manner such that there is no interaction that would substantially impair the desired pharmaceutical efficacy.

Dose of the pharmaceutical compositions of the invention will vary depending on the subject and upon particular route of administration used. By way of an example only, an overall dose range of from about 1 microgram to about 300 micrograms or 0.1 to 100 mg/kg/day is contemplated for human use. Pharmaceutical compositions of the present invention can also be administered to a subject according to a variety of other, well-characterized protocols. Desired time intervals for delivery of multiple doses of a particular composition can be determined by one of ordinary skill in the art employing no more than routine experimentation.

The peptides of the invention may also be administered per se (neat) or in the form of a pharmaceutically acceptable salt. When used in medicine, the salts should be pharmaceutically acceptable but non-pharmaceutically acceptable salts may conveniently be used to prepare pharmaceutically acceptable salts thereof and are not

excluded from the scope of this invention. Such pharmaceutically acceptable salts include, but are not limited to, those prepared from the following acids: hydrochloric, hydrobromic, sulphuric, nitric, phosphoric, maleic, acetic, salicylic, p-toluene-sulfonic, tartaric, citric, methanesulphonic, formic, malonic, succinic, naphthalene-2-sulfonic, and benzenesulphonic. Also, pharmaceutically acceptable salts can be prepared as alkaline metal or alkaline earth salts, such as sodium, potassium or calcium salts of the carboxylic acid group.

The compositions include those suitable for oral, rectal, topical, nasal, ophthalmic or parenteral administration, all of which may be used as routes of administration using the materials of the present invention. Other suitable routes of administration include intrathecal administration directly into spinal fluid (CSF), direct injection onto an arterial surface and intraparenchymal injection directly into targeted areas of an organ. Compositions suitable for parenteral administration are preferred. The term "parenteral" includes subcutaneous injections, intravenous, intramuscular, intrasternal injection or infusion techniques.

The compositions may conveniently be presented in unit dosage form and may be prepared by any of the methods well known in the art of pharmacy. All methods include the step of bringing the active ingredients of the invention into association with a carrier which constitutes one or more accessory ingredients.

Compositions of the present invention suitable for oral administration may be presented as discrete units such as capsules, cachets, tablets or lozenges, each containing a predetermined amount of the peptides of the invention or as a suspension in an aqueous liquor or non-aqueous liquid such as a syrup, an elixir, or an emulsion.

Preferred compositions suitable for parenteral administration conveniently comprise a sterile aqueous preparation of peptides of the invention which is preferably isotonic with the blood of the recipient. This aqueous preparation may be formulated according to known methods using those suitable dispersing or wetting agents and suspending agents. The sterile injectable preparation may also be a sterile injectable solution or suspension in a non-toxic parenterally-acceptable diluent or solvent, for example as a solution in 1,3-butane diol. Among the acceptable vehicles and solvents that may be employed are water, Ringer's solution and isotonic sodium chloride solution. In addition, sterile, fixed oils are conventionally employed as a solvent or suspending medium. For this purpose any bland fixed oil may be employed including

synthetic mono-or diglycerides. In addition, fatty acids such as oleic acid find use in the preparation of injectables.

The following non-limiting examples are illustrative of the invention.

General Comments

The following laboratory procedures were used in the examples below.

1. Fusion Protein Expression and Purification

GST-fusion proteins were expressed in either DH5α or BL21 bacterial strains. Cultures with an $OD_{600}$ of 0.2 were induced for three hours with isopropyl β-D-thiogalactopyranoside (IPTG). Bacteria were harvested by centrifugation and resuspended in 10 mL of NETN buffer which contains 20 mM tris(hydroxymethyl)aminomethane (Tris), pH 8.0, 100 mM NaCl, 1 nM ethylenediamine tetraacetic acid (EDTA), 0.5% NP-40, and 2 mM phenylmethylsulfonyl fluoride (PMSF). The bacterial cells were lysed by sonication. Affinity purification using glutathione-sepharose beads was carried out according to protocols provided by the manufacturer (Pharmacia Biotech Inc., Uppsala, Sweden.

Fusion proteins can also be prepared using other fusion protein systems known in the art including those set forth in U.S. Patents 5,270,170 and 5,498,530, both of which are herein incorporated by reference.

2. Library Construction

The random 15-mer library was constructed as described in detail by P. Schatz et al., *Meth. Enzymol.*, **267**:171-191 (1996), which is herein incorporated by reference, using an oligonucleotide with a degenerate region of 15 codons in the form of NNK, where N denotes an equimolar mix of all four bases and K denotes a mix of G or T. The library consisted of $1.3 \times 10^{10}$ independent recombinants. The amplified library were stored at -80°C in HEK buffer containing 35 mM HEPES pH 7.5, 0.1 mM EDTA, and 50 mM KCl.

Random peptide libraries may also be constructed using other DNA binding protein/specific binding site systems such as those disclosed in U.S. Patent Nos. 5,498,530 and 5,270,170, each of which is herein incorporated by reference.

3. Construction of maltose binding protein fusions

Nucleotide sequences encoding appropriate peptides were cloned into pELM3 (P. Schatz et al., *Meth. Enzymol.*, **267**:171-191 (1996)). This allows expression of the corresponding maltose binding protein/peptide fusion. The procedure for expression

of maltose binding proteins was identical to that for GST fusions except that the LB medium was supplemented with 2% glucose.

4. Affinity Panning

A 2 ml aliquot of thawed bacterial cells in HEK was added to 6 ml of lysis buffer 25 mM HEPES pH 7.5, 0.07 mM EDTA, 8.3% glycerol, 1.25 mg/ml bovine serum albumin (BSA), 0.83 mM DTT, 0.2 mM PMSF. The bacteria were lysed for 2 to 4 mm on ice by the addition of 0.15 ml 10 mg/ml lysozyme (Boehringer Mannheim, Indianapolis, IN) and then 2 ml of 20% lactose and 0.25 ml of 2 M KCl were added. The supernatant was obtained after a 15 mm centrifugation at 27,000 x g. To initiate panning, 12 wells of a 96-well plate were first coated with GST-fusion proteins (10 μg protein per well) at 4°C for 1 hour. The wells were then blocked with 1% BSA in phosphate-saline buffer (PBS) at pH 7.4. After precoating, 250 μl of the supernatant was added to each of precoated wells. After gentle agitation for 1 hour at 4°C, the unbound material was recovered and the wells were then washed with a series of solutions: 5 times with HEK buffer supplemented with 0.2M lactose and 1% BSA, twice with HEK supplemented with 0.2 M lactose, and twice with HEK at 4°C. The bound plasmids were eluted with 35 mM HEPES, pH 7.5, 0.1 mM EDTA, 200 mM KCl, 1 mM IPTG for 30 mm at room temperature. The eluted DNA was precipitated with isopropanol and amplified by electrotransformation. This pool of bacterial transformants were used in subsequent rounds of panning.

The panning procedure was monitored by two parameters: recovery and enrichment. Recovery was calculated by subtracting the number of plasmids bound to receptor/BSA-coated wells by number of plasmids bound to BSA-coated wells. The enrichment at each round of panning was the ratio of recovered plasmids from receptor coated wells to those recovered from BSA coated wells. The details of one affinity panning using PDZ3 of PSD-95 is shown:

| Round No. | Input | Output | Recovery | Enrichment |
|---|---|---|---|---|
| 1 | $6.0 \times 10^{9}$ | $1.72 \times 10^{5}$ | $2.9 \times 10^{-5}$ | - |
| 2 | $3.2 \times 10^{9}$ | $1.4 \times 10^{5}$ | $4.4 \times 10^{-5}$ | 3 |
| 3 | $1.2 \times 10^{8}$ | $1.1 \times 10^{6}$ | $5.9 \times 10^{-3}$ | 270 |
| 4 | $8.4 \times 10^{7}$ | $4.8 \times 10^{6}$ | $5.7 \times 10^{-2}$ | 1,700 |

5. ELISA

After three to four rounds of affinity panning, individual colonies were randomly selected. Overnight cultures from single colonies were diluted 1:10 in 3 ml of LB ampicillin (100 μg/ml) and grown 1 hour at 37°C. The expression of the LacI-peptide fusions was induced by the addition of arabinose to 0.2% for 3 hours. After induction, the cells were pelleted by centrifugation and lysed as described above in 1 ml of lysis buffer plus lysozyme. The clarified lysates were used immediately for ELISA or stored at -70°C. To prepare ELISA, 96-well plates were first coated with GST- fusion proteins (0.2 μg protein per well) of nNOS, PSD-95, or disheveled PDZ domain at 4°C for 1 hour. The wells were then blocked with 1% BSA in phosphate-saline buffer (PBS) at pH 7.4. After precoating, the wells were washed three times with PBS supplemented with 0.05% Tween-20 (PBT). To initiate the binding, 100 μl of 1:10 diluted lysate was added to each well. After 30 minutes at 4°C, the plate was washed four times with PBT. The binding of LacI-peptide was detected using rabbit anti-Lac I antibody. After 4 washes with PBT, the plate was developed by adding alkaline phosphatase-conjugated goat anti-rabbit antibody (GIBCO-BRL, Gaithersburg, MD) in PBS/0.1% BSA (100 μl per well for 1 hour at 25°C) followed by a 6 mm treatment with p-nitrophenyl phosphate (4 mg/ml) in 1 M diethanolamine hydrochloride, pH 9.8/0.24 mM $MgCl_2$ (200 μl per well). Binding was quantified by monitoring optical density (O.D.) at 405 nm on an E-max plate reader (Molecular Devices Inc., Melno Park, CA). The negative controls were wells coated with control GST fusion or as otherwise indicated. All experiments were repeated at least once with similar results.

ELISAs for maltose binding fusion proteins were performed as described above with a few modifications. 100 μl of a 1:50 dilution of crude lysate was added to each well. All buffers were the same but were supplemented with 1 mM maltose to minimize oligomerization of maltose binding protein fusions (G. Richarme, *Biochemical and Biophysical Research Communications*, **105**:476-481 (1982)). Interaction of maltose binding protein fusion proteins with immobilized GST-fusion proteins was monitored by rabbit anti-maltose binding protein antibody (1:10,000 dilution, New England Biolabs, Inc., Beverly, MA).

6. Peptide-PDZ binding

To determine the affinity of peptide-PDZ interactions, monomeric maltose binding protein fusions of peptides were purified by amylose affinity columns according to a protocol provided by the manufacturer (New England Biolabs, Inc., Beverly, MA). Protein concentration was determined by the Bradford assay (BioRad, Richmond, CA) using BSA as standard. The effective concentration, i.e., $EC_{50}$ was determined by dose dependent ELISA tests. GST fusion was bound at 0.05 μg per well. The maltose binding protein fusions were incubated after being serially diluted (1:5) starting at 15 μM. The data were fit with the Hill equation ($O.D.405=O.D._{405Max}/1+ \{EC_{50}/[x]\}$ n). A non-linear least square algorithm was used.

7. Yeast Two Hybrid Analysis

Yeast Y187 cells were co-transformed with expression vectors encoding various Gal4 DNA binding domain-nNOS fusions and the Gal4 activation domain fused to PSD-93 (amino acids 116-421). Each transformation mixture was plated on synthetic dextrose plates lacking tryptophan and leucine. Interaction was measured by the liquid culture β-galactosidase assay as described (S. Fields et al., *Nature*, **340**:245-246 (1989); and Song, 1989; Clonetech, Palo Alto, CA)). Values are representative of duplicate experiments.

8. Fusion Protein Affinity Chromatography

Rat whole brain was homogenized in 10 volumes (w/v) tris-HCl buffer pH 7.4 and centrifuged at 32,000 x g for 20 minutes. Membranes were solubilized for 2 hours at 4°C in buffer containing 200 mM NaCl and 1% Triton X-l00 and insoluble material pelleted by centrifugation at 100,000 x g for 30 minutes. Extracts were incubated with control amylose beads or amylose beads saturated with maltose-binding fusion proteins as indicated. Samples were loaded into disposable columns, which were washed with 50 volumes of buffer containing 1% Triton X-100 + 300 mM NaCl. Retained proteins were eluted with 150 μl of loading buffer and were resolved by SDS / PAGE. Blots were hybridized with a monoclonal antibody to nNOS (Transduction Labs, Lexington, KY).

Example 1 - Construction of a random C-terminal random C-terminal peptide library

Peptide binding and x-ray crystallographic studies of PSD-95 indicate that specificity of the peptide-PDZ interaction is primarily determined by the final 4

residues of the peptide ligand (D. Doyle et al., *Cell*, **85**:1067-1076 (1996); E. Kim et al., *Nature*, **378**:85-88 (1995); H. Kornau et al., *Science*, **269**:1737-1740 (1995); B. Muller et al., *Neuron*, **17**:255-265 (1996); M. Niethammer et al., *J. Neurosci.*, **16**:2157-2163 (1996)). To determine optimal peptide binding ligands for other PDZ domains, we constructed a fusion protein library that contains 15 randomized residues at the C-terminus. In this library, a degenerate oligonucleotide encoding the random peptides is fused to the end of the *E.coli* lac repressor (M. Cull et al., *Proc. Natl. Sci. USA*, **89**:1865-1869 (1992)), which is herein incorporated by reference. Following expression1 the Lac repressor protein binds to the lac operator sequence on the same plasmid linking each randomized 15-mer peptide to the plasmid encoding that peptide (Figure 1). This linkage allows repeated rounds of selection for specific peptide ligands in the population by affinity purification of peptide-repressor-plasmid complexes (see the experimental procedures set forth above).

*In vitro* selection of optimal binding peptides for PDZ domains

A random 15-mer peptide library using the third PDZ (PDZ3) domain of PSD-95 was screened according to the following steps. Step I. A pool of oligonucleotides encoding 15 random amino acids ($X_{15}$) was cloned in frame C-terminal to *lac I*. Protein expression from each plasmid of the library yields a Lac I fusion with a distinct peptide sequence. The recombinant Lac I binds the *lac 0* sites present on the same plasmid yielding Lac I-plasmid complexes that are purified from the *E.coli*. Step II. Affinity panning selects peptides that interact with target receptor1 e.g., PDZ domain. Step III. The bound plasmid DNA can be specifically recovered by addition of IPTG. Step IV. The recovered plasmids are retransformed, amplified, and used for subsequent rounds of panning.

In PSD-95, PDZ1 and PDZ2 domains interact with the C- terminal four amino acids found in Shaker potassium channels and NMDA receptor subunits (H. Kornau et al., *Science*, **269**:1737-1740 (1995); E. Kim et al., *Nature*, **378**:85-88 (1995)), which have a shared consensus of E-(T/S)-X-V-COOH. PDZ3 binds to an identical sequence (D. Doyle et al., *Cell*, **85**:1067-1076 (1996)). A PDZ3 fusion protein was constructed by linking amino acids 302-402 of PSD-95 to the C-terminus of glutathione S-transferase (GST). The purified protein was incubated with a 15-mer *lac I* library with a complexity of $1.3 \times 10^{10}$. After 4 rounds of panning selection, a 1,700-fold enrichment of interacting peptides was achieved (see Experimental

procedures). At this stage, individual clones were randomly selected and subjected to ELISA analysis (Figure 2A).

Briefly, crude bacterial lysates from individual clones (horizontal axis of Figure 2A) selected through four rounds of panning were prepared (see Experimental procedures). Association of Lac 1-peptide fusion with GST-PDZ3 was determined by ELISA. Dashed bars indicate wells coated with BSA only; gray bars: GST-$NAB_{HERG}$ + BSA; open bars: GST-nNOS-PDZ + BSA; closed bars: GST-PDZ3 + BSA. GST-$NAB_{HERG}$ is a fusion protein containing amino acids 1-135 from *HERG* potassium channel which has no homology with PDZ domain (X. Li et al., *J. Biol. Chem.*, **272**(2):705-708 (1997)). All ELISA experiments in this figure and subsequent figures have been repeated at least once with similar results.

Enriched clones were divided into two classes. One class, such as PD-301, PD-302, and PD-304, interacted with both GST control and GST-PDZ3 fusion (Figure 2A), suggesting that the corresponding peptides interact with GST. The other class of clones, including PD-312, PD-314, and PD-315, bound selectively to GST-PDZ3. Affinity of interaction ($EC_{50}$) was 2 to 100 nM as determined by quantitative ELISA as set forth above.

To determine the binding specificity1 purified recombinant PDZ fusion proteins of nNOS (amino acids 1-150, D. Bredt et al., *Nature*, **351**:714-718 (1991)) and disheveled (amino acids 146-226; J. Klingensmith et al., *Genes Dev.*, **8**:118-130 (1994)) were also tested for peptide-binding. Under the same conditions, the PDZ3-positive clones failed to interact with the PDZ domain of nNOS (Figure 2A) or with the PDZ domain of disheveled. Plasmids encoding PDZ3-specific clones were sequenced.

An alignment of the deduced amino acid sequences is shown (Figure 2B). Indeed, most of the interacting peptides closely resemble the peptide sequence at the C-terminus of Shaker-like potassium channels and NMDA receptor subunits, with a consensus of E-(T/S)-X-V-COOH.

Identification of novel peptides interacting with PDZ domain of nNOS

To determine optimal peptide ligands for the nNOS PDZ domain, a recombinant GST fusion protein corresponding to the coding sequence of amino acids 1 to 150 of nNOS (nNOS-PDZ) was used for peptide selection. After four rounds of panning1 a 2,300-fold enrichment was achieved. Individual GST-nNOS-PDZ

specific clones were identified by ELISA (Figure 3A). It was discovered that 95 out of 150 clones specifically interacted with nNOS-PDZ but not with the control GST fusion protein. Binding affinity of these peptides to immobilized nNOS-PDZ ($EC_{50}$) was 8 to 100 nM. Plasmids from these nNOS specific clones were sequenced. The deduced amino acid sequences of 95 independent clones were aligned via their C-termini (Figures 3B and 3C).

An analysis of amino acid abundance at each position indicates that valine again is strongly preferred (89%) at the 0 position (Figures 4A-4I). At the -1 position, there is no obvious preference. Fifteen of the twenty amino acids were found - amino acids D, E, H, K and N were not present. In contrast to the PDZ3 consensus, aspartate at the -2 position was present in 81% of all nNOS-PDZ binding peptides. At the -3 position, glycine is significantly preferred. Considering that glycine was used as a part of the linker that separates Lac I from the random peptide (Figure 1), this bias was appropriately corrected. The corrected glycine abundance is 47% at the -3 position. From position -4 to position -8, no obvious amino acid preference was observed (Figures 4A-4I). Based on the amino acid abundance at each position, the optimal sequence for a nNOS binding peptide (NBP) is g-D-X-V-COOH.

SPECIFICITY OF NBP BINDING TO NNOS-PDZ

Figures 5A- D show that NBP's bind specifically to nNOS PDZ and native nNOS protein from rat brain.

The *in vitro* peptide selection suggests that PDZ3 of PSD-95 and the nNOS-PDZ, despite a shared preference for valine at the 0 position1 have distinct binding specificity. To directly test this, we performed ELISA as set forth above and found that 36 randomly chosen NBPs failed to bind to PDZ3 of PSD-95 (Figure 5A) or to the PDZ domain of disheveled. Based on the peptide-PDZ3 crystal structure (D. Doyle et al., *Cell*, **85**:1067-1076 (1996)), it is known that the side-chain of His372 of PSD-95 forms a critical sequence specific hydrogen bond with the T at the -2 position of the bound peptide. Interestingly, the amino acid at the corresponding position of nNOS-PDZ is Y77, consistent with the idea that substitution of H to Y at this position converts the -2 position peptide preference from T to D. Also in agreement with this notion, the corresponding residue of the disheveled PDZ is N. Amino acid sequence comparison of a number of PDZ domains present in Genbank shows that the residue after the H or Y is also conserved (nNOS is Y-D, PDZ3 is H-E). To determine

whether the Y77 of nNOS is critical we mutated Y77D78 to H77E78. This mutant, nNOS-PDZHE, lost its ability to bind D-X-V peptides and gained the ability to bind T-X-V peptides (Figure 5B).

To evaluate the specificity of the NBP-nNOS interactions, we mutated the D at the -2 position of the NBP-123 (LDRLRNRVHGDAV-COOH, $EC_{50}$=40 nM) peptide to A, L, Q, R, S, T, and V. Peptides with these amino acid substitutions failed to interact with nNOS-PDZ (Figure 5C). To test whether NBPs bind to native nNOS protein, we generated an affinity column linking NBP-123 to an agarose matrik (see the experimental procedures set forth above). We found that nNOS protein in crude rat brain homogenates adhered to the NBP-123 matrix. In contrast, nNOS did not bind to an analogous column in which the -2 D residue of NBP-123 was mutated to T (Figure 5D).

The nNOS-PDZ Domain Has Unique Structural Feature

Previous studies have shown that the N-terminal domain of nNOS (amino acids 1-150) binds to the PDZ domain of (1-syntrophin and to the second PDZ domains of PSD-95 and PSD-93 (J. Brenman et al., *Cell*, **84**:757-767 (1996)). Although amino acids 16 to 100 of nNOS define the consensus PDZ domain, binding studies have shown that fusions containing amino acids 1 to 100 of nNOS do not bind to the PDZ domain of either OLl-syntrophin or PSD-93 (J. Brenman et al., *Cell*, **84**:757-767 (1996)). To test whether the peptide binding property of the nNOS-PDZ is confined to the typical consensus, we tested whether any of five randomly selected NBPs interact with a fusion protein containing nNOS 1-100. We found that all 5 NBPs bind to nNOS (1-150) but not to nNOS (1-100).

To determine the minimal functional structure for nNOS-PDZ to bind NBPs and PSD-93, we generated a panel of six fusion proteins that express various regions of the N-terminus of nNOS (Figure 6). We first evaluated binding of these constructs to the PDZ repeats in PSD-93 using the yeast two-hybrid analysis. Binding to PSD-93 required amino acids 16-130 of nNOS; truncations on either side of this core nNOS- PDZ eliminate the interaction. Similarly, all NBPs required amino acids 16-130 for binding as tested by ELISA (Figure 6). These studies indicate that the functional nNOS-PDZ requires additional amino acids beyond the conserved consensus and indicate that both peptide-PDZ and PDZ-PDZ interactions of nNOS likely require a similar tertiary structure.

Candidate proteins that interact with nNOS

Identification of the ligand binding consensus of nNOS- PDZ allows an electronic search for potential nNOS interacting proteins present in the protein databases. A pre-release version of the XREFPatScan software, written in the perl programming language was used to find all occurrences of the D-X-V pattern at the carboxy-terminus of protein sequences in the non-redundant protein database (nr, 11 Nov 1996) maintained at the National Center for Biotechnology Information http://www.ncbi.nlm.nih.gov). This sequence pattern scan has revealed 484 matches in the database. Interestingly, this list of potential binding partners includes both glutamate and melatonin receptors, which are well known to influence nNOS activity. See Figures 8A-8R for more detailed results of the PDZ scan of the database.

Another suitable software package is the SASP package available from GCG (Genetics Computer Group, University Research Park, Madison WI).

In summation, we have employed a powerful genetic strategy to identify C-terminal peptide ligands for the nNOS PDZ domain. This strategy takes advantage of the strong protein-DNA association between the lac repressor and the lac operator sequence. This interaction is used to obtain a highly complex library of expressed peptides each bound to the plasmid that encodes them. By simply panning for peptide binding and then sequencing the corresponding plasmids, we were able to rapidly determine optimal binding partners for the nNOS-PDZ. Identified peptides bind potently to nNOS with binding affinities ($EC_{50}$) in the 8-100 nM range, similar to the affinity between the NMDA receptor and PDZ domain of PSD-95 (B. Muller et al., *Neuron*, **17**:255-265 (1996)). These peptide sequences are likely to be physiologically relevant because a similar panning procedure yielded the known peptide ligands for PDZ3 of PSD-95.

The consensus peptide binding sequence for the nNOS-PDZ is D-X-V, which contrasts with the E-(T/S)-x-V found for PDZs of PSD-95 (D. Doyle et al., *Cell*, **85**:1067-1076 (1996); E. Kim et al., *Nature*, **378**:85-88 (1995); H. Kornau et al., *Science*, **269**:1737-1740 (1995); B. Muller et al., *Neuron*, **17**:255-265 (1996); M. Niethammer et al., *J. Neurosci.*, **16**:2157-2163 (1996)). Analysis of the crystal structure of peptide-bound PDZ3 suggests rational explanations for these alternate specificity (D. Doyle et al., *Cell*, **85**:1067-1076 (1996)). Similar preference of the two domains for terminal valine is expected because the critical residues in the

carboxylate binding loop of PDZ3, including the GLGF tetrapeptide, are precisely conserved in nNOS-PDZ. While the carboxylate loop of PSD-95 binds most potently to peptides with C-terminal valine, other terminal hydrophobic amino acids are permitted. Such degeneracy was also found in some nNOS binding peptides, e.g., NBP-14 (Figures 3B and 3C). Inwardly rectifying potassium channel subunits of class 2.0 terminate with S-X-I and these channels also bind to PSD-95. In addition the -2 serine of Kir 2.3 serves as a potent substrate for protein kinase A and this phosphorylation event regulates binding of the channel to PSD-95 (N. Cohen *Neuron*, **17**:759-767 (1996)).

Specificity of PDZ3 for T/S at the peptide -2 position is mediated by hydrogen bonding of the hydroxyl of the T/S with the N-3 nitrogen of H372 of PDZ3 (D. Doyle et al., *Cell*, **85**:1067-1076 (1996)). The corresponding residue in nNOS is Y77. The greater electrophilic character of Y compared to H may explain the preference of the nNOS PDZ for the acidic amino acid D at peptide position -2. Accordingly, mutation of Y77D78 of nNOS to H77E78 changes the binding specificity from DXV to TXV. Interesting, the Y77 position is not generally conserved in other orphan PDZ domains and this single residue may allow for much of the diverse peptide ligand specificity at the -2 position.

These studies emphasize that the nNOS PDZ domain has unique structural features. The consensus PDZ domain contains 80 amino acids, and PDZ3 of PSD-95 was functionally active as a 101 amino acid polypeptide (D. Doyle et al., *Cell*, **85**:1067-1076 (1996)). By contrast, a functional nNOS PDZ domain requires an additional 30 amino acids C-terminal to the identified consensus. We wondered whether the smaller nNOS constructs, such as nNOS 1-100, were inactive due to a non-specific problem with polypeptide folding. However, circular dichroism (CD) analysis indicated a predicted high degree of secondary structure for nNOS 1-100 consisting of ~X% of α-helix and ~Y% β-strand. This is similar to the composition of α-helix and β-strand found in PDZ3 structure of PSD-95. Furthermore nNOS 1-100 showed thermal stability to 42°C which is comparable to the thermal stability of a functionally active PDZ domain of FAP. Therefore, we believe that the functional nNOS PDZ has a structure somewhat larger than that of other PDZ domains. By using our genetic peptide selection strategy, it will be possible to determine whether other PDZ domains are also larger than the presently identified consensus. See K.

Christopherson et al., *J. Clin. Invest.*, **100**:2424-2429 (1997); and N. Stricker et al., *Nat. Biotechnol.*, **15**:336-342 (1997), the disclosures of which are hereby incorporated by reference.

In addition to interacting with peptide ligands, the PDZ domain of nNOS associates with other PDZ domains, including the PDZ domain of (1-syntrophin and the second PDZ of PSD-95 and PSD-93. Three dimensional structure of a PDZ/PDZ heterodimer is not yet available, but our data suggest the PDZ / PDZ binding interface overlaps with the peptide recognition sequences. Thus, deletions of nNOS PDZ that abolish peptide binding also eliminate binding to (1-syntrophin and PSD-93. Crystallography of PDZ3 of dlg showed that the PDZ domain forms a dimer in which the surface of the peptide-binding domain of one PDZ subunit interacts with residues in (-strands from the other subunit (J. Cabral et al., *Nature*, **382**:649-652 (1996)). This binding topology of PDZ domains may explain why the SXV peptide of the NMDA receptor 2B potently blocks nNOS binding to PSD-95 (J. Brenman et al., *Cell*, **84**:757-767 (1996)). Proteins containing the DXV nNOS interacting domain may also disrupt interaction of nNOS with PDZ proteins. This may explain the paradoxical situation that (l-syntrophin, but not nNOS, is present at the sarcolemma in patients with Becker muscular dystrophy (D. Chao et al., *Journal of Experimental Medicine*, **184**:609-618 (1996)). Perhaps, in the myofibers of these patients, the nNOS PDZ is occupied by a protein with a C-terminal D-X-V and is unable to bind to OLl-syntrophin.

The disclosed genetic selection strategy will help identify peptide ligands for the 100s of orphan PDZ domains that have been sequenced. After isolating high affinity peptides, protein data base analysis may suggest candidate physiological binding partners. Our search with the terminal DXV consensus for nNOS yielded several attractive candidates including melatonin receptor la (U14108) and an alternatively spliced form of GluR6 (X66117). Though nNOS is best activated by calcium influx through NMDA receptors (J. Garthwaite et al., *Nature*, **336**:385-388 (1988)), there is also abundant literature showing that nNOS activity can be regulated by melatonin (D. Vesely, *Mol. Cell Biochem.*, **35**:55-58 (1981)) and by non-NMDA type glutamate receptors (J. Garthwaite et al., *Annu. Rev. Physiol.*, **57**:683-706 (1995)). Our data suggest that physical association of nNOS with GluR6 and with melatonin receptors may participate in this functional coupling.

The invention has been described with reference to preferred embodiments thereof. However, it will be appreciated that those skilled in the art, upon consideration of this disclosure, may make modifications and improvements within the spirit and scope of the invention as set forth in the following claims.

What is claimed is:

1. A peptide of at least 3 amino acids comprising the sequence D-X-V-COOH wherein D=Aspartic acid, X=any amino acid and V=Valine.

2. An isolated nucleic acid encoding the peptide of claim 1.

3. A method for determining the identity of proteins which interact with a protein binding domain (orphan protein domain) of a first protein (Protein Interaction Network (PIN)) comprising:

screening a random peptide library comprising transformed host cells, each of which contains a plasmid that comprises a *lacO* binding site and encodes a fusion protein comprising a Lac repressor DNA binding protein fused to a peptide, wherein each transformed host cell differs from one another with respect to the peptide in said fusion protein, said screening comprising lysing the host cells under conditions that the fusion protein remains bound to the plasmid at the *lacO* binding site, contacting the fusion proteins of the random peptide library with a protein binding domain (orphan protein domain) under conditions conducive to specific peptide-protein binding domain (orphan protein domain) binding;

isolating the plasmid that encodes a peptide that binds to the protein binding domain (orphan protein domain);

sequencing the plasmid to obtain the sequence of the peptide that binds to the protein binding domain (orphan protein domain); and

searching the available nucleic acid and protein sequence databases to identify proteins which comprise the sequence of the peptide which binds to the protein binding domain (orphan protein domain)

4. The method of claim 3, further comprising the step of: assembling the PINS from different orphan protein domains into an electronic databank that can be searched with a the sequence of a protein domain (orphan protein domain) of interest.

5. A method of treating a neurodegenerative disease, motility disorder or muscular dystrophy in a human or animal comprising administering to a patient in need thereof an effective amount of the peptide of claim 1.

6. The peptide of claim 1, wherein said peptide comprises at least 5 amino acids.

7. The peptide of claim 1, wherein said peptide comprises at least 10 amino acids.

8. The peptide of claim 1, wherein said peptide comprises at least 15 amino acids.

9. A peptide ligand detection system comprising:

a) a random peptide library comprising a recombinant DNA vector encoding a DNA binding protein that specifically binds a DNA sequence on the vector, the DNA binding protein comprising a covalently linked sequence encoding a random peptide sufficient for the vector to encode at least about $10^6$ different fusion proteins each of which is capable of specifically binding the DNA sequence on the vector; and

b) an orphan protein domain sequence immobilized on a solid support capable of specifically binding the random peptide of the DNA binding protein.

10. The peptide ligand detection system of claim 9 further comprising an inducer molecule capable of specifically binding the DNA binding protein sufficient to release the recombinant DNA vector from the immobilized orphan protein domain sequence.

11. The peptide ligand detection system of claim 9 wherein the DNA binding protein comprises a prokaryotic repressor protein sequence and the DNA sequence bound by the DNA binding protein is a prokaryotic operator sequence.

12. The peptide ligand detection system of claim 11 wherein the prokaryotic repressor protein sequence is a lac repressor or a fragment thereof capable of specifically binding the DNA sequence on the vector.

13. The peptide ligand detection system of claim 11 wherein the prokaryotic operator sequence is lac O or a fragment thereof capable of being specifically bound by the prokaryotic repressor protein sequence.

14. The peptide ligand detection system of claim 10 wherein the inducer molecule is isopropylthio-β-D-galactoside (IPTG).

15. The peptide ligand detection system of claim 11 wherein the prokaryotic repressor protein sequence and the random peptide sequence are spaced by a peptide linker sequence encoded by nucleic acid sequence comprising -G-G-G-.

16. A peptide ligand detected by the ligand detection system of claim 1 having a binding affinity ($EC_{50}$) for the orphan protein domain of between about 0.5 to 500 nM.

17. A peptide ligand comprising between about 3 and 50 amino acids comprising an amino acid sequence consisting of D-X-V-COOH, wherein the peptide ligand has a binding affinity ($EC_{50}$) for an orphan protein domain of between about 0.5 to 500 nM.

18. The peptide ligand of claim 17, wherein the orphan protein domain is a PDZ domain.

19. The peptide ligand of claim 18, wherein the PDZ domain is obtained from a protein selected from the group consisting of nitric oxide synthase (nNOS), post-synaptic density protein (PSD-95/SAP-90), post-synaptic density protein (PSD-93), epithelial tight-junction protein zona occludens-1 (ZO1), N-methyl-D-aspartate (NMDA) type glutamate receptor, Shaker-type potassium channel subunit, and 1-syntrophin.

20. A therapeutic composition comprising the peptide ligand of claim 18.

21. An isolated nucleic acid encoding the peptide ligand of claim 18.

22. A DNA vector comprising the isolated nucleic acid of claim 21.

23. A method of detecting a peptide ligand capable of specifically binding an orphan protein domain of a protein, the method comprising:

a) lysing transformed cells comprising a random peptide library comprising a recombinant DNA vector encoding a DNA binding protein that specifically binds a DNA sequence on the vector, the DNA binding protein comprising a covalently linked sequence encoding a random peptide sufficient for the vector to encode at least $10^6$ different fusion proteins each of which is capable of specifically binding the DNA sequence on the vector, wherein the lysing is under conditions such that the DNA binding protein comprising the random peptide remains bound to the recombinant DNA vector,

b) contacting the fusion proteins of the random peptide library to an immobilized orphan protein domain under conditions conducive to specific peptide-orphan protein domain binding; and

c) isolating a recombinant DNA vector encoding a fusion protein that specifically binds to the orphan protein domain.

24. The method of claim 23 further comprising the steps of transforming a host cell with the recombinant DNA vector obtained in step c), repeating steps a), b), and c) with the host cell, and isolating a selected recombinant DNA vector.

25. The method of claim 24 further comprising determining the amino acid sequence of the random peptide encoded by the selected recombinant DNA vector.

26. The method of claim 25 further comprising searching a protein sequence database to identify an orphan protein domain in the database comprising the random
peptide.

27. The method of claim 26 further comprising assembling a protein interaction network (PIN) sufficient to correlate a plurality of random peptide sequences to the orphan protein domain.

28. The method of claim 27 further comprising assembling a super protein interaction network (SPINS) comprising a plurality of protein interaction networks (PINs) sufficient to serve as an electronic extension database for the protein sequence database.

29. The method of claim 26 wherein the orphan protein domain in the database is any one of the orphan protein domains (protein modules) shown in Figure 7.

30. A method of detecting a peptide ligand capable of specifically binding an orphan protein domain of interest, the method comprising searching a super protein interaction network (SPINS) with an amino acid sequence comprising an orphan protein domain of interest, and identifying the peptide ligand capable of specifically binding the orphan protein domain of interest.

31. The method of claim 30, wherein the peptide ligand is obtained from a random peptide library.

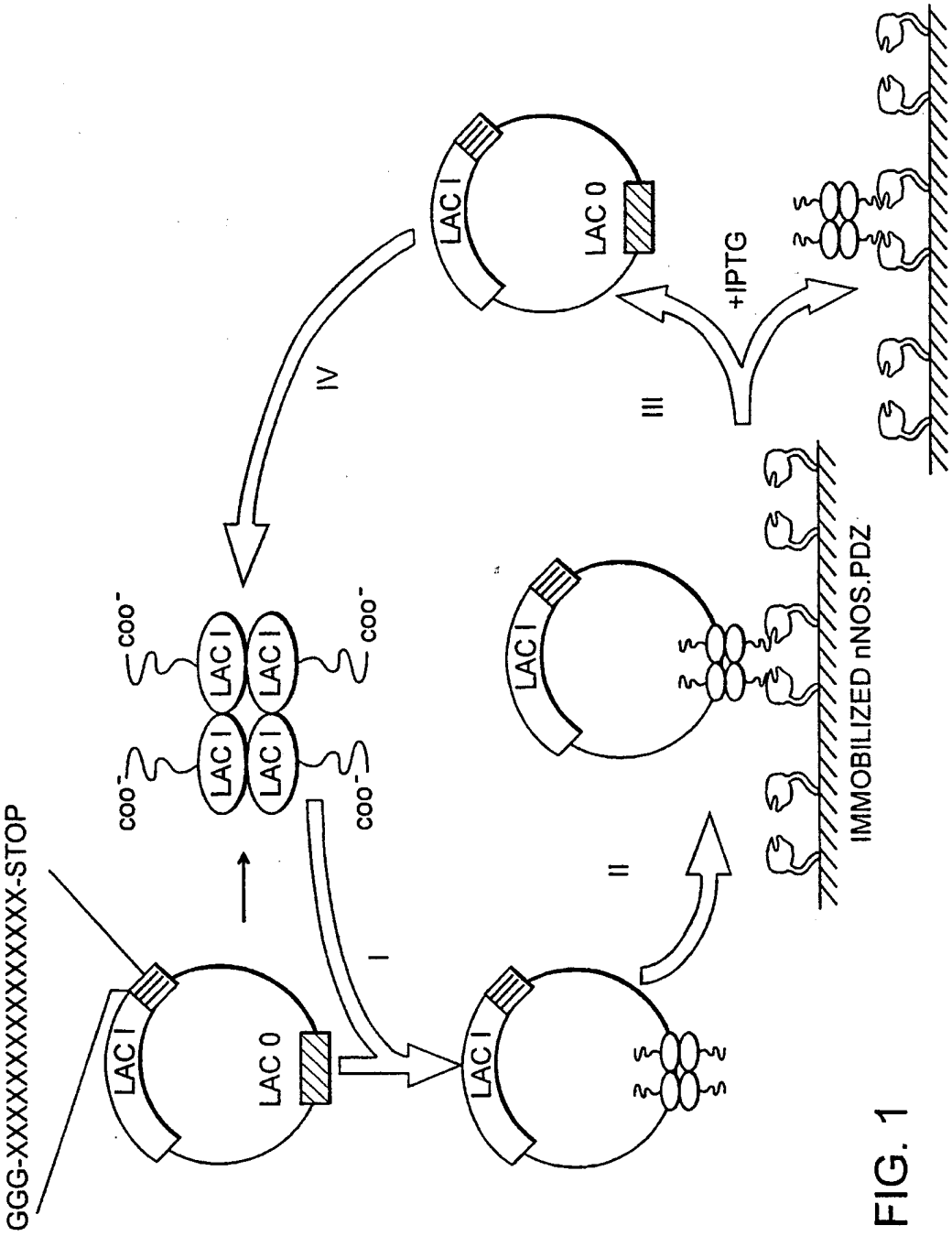
GGG-XXXXXXXXXXXX-STOP
LAC I
LAC 0
I
II
III
IV
+IPTG
IMMOBILIZED nNOS.PDZ
coo-

FIG. 1

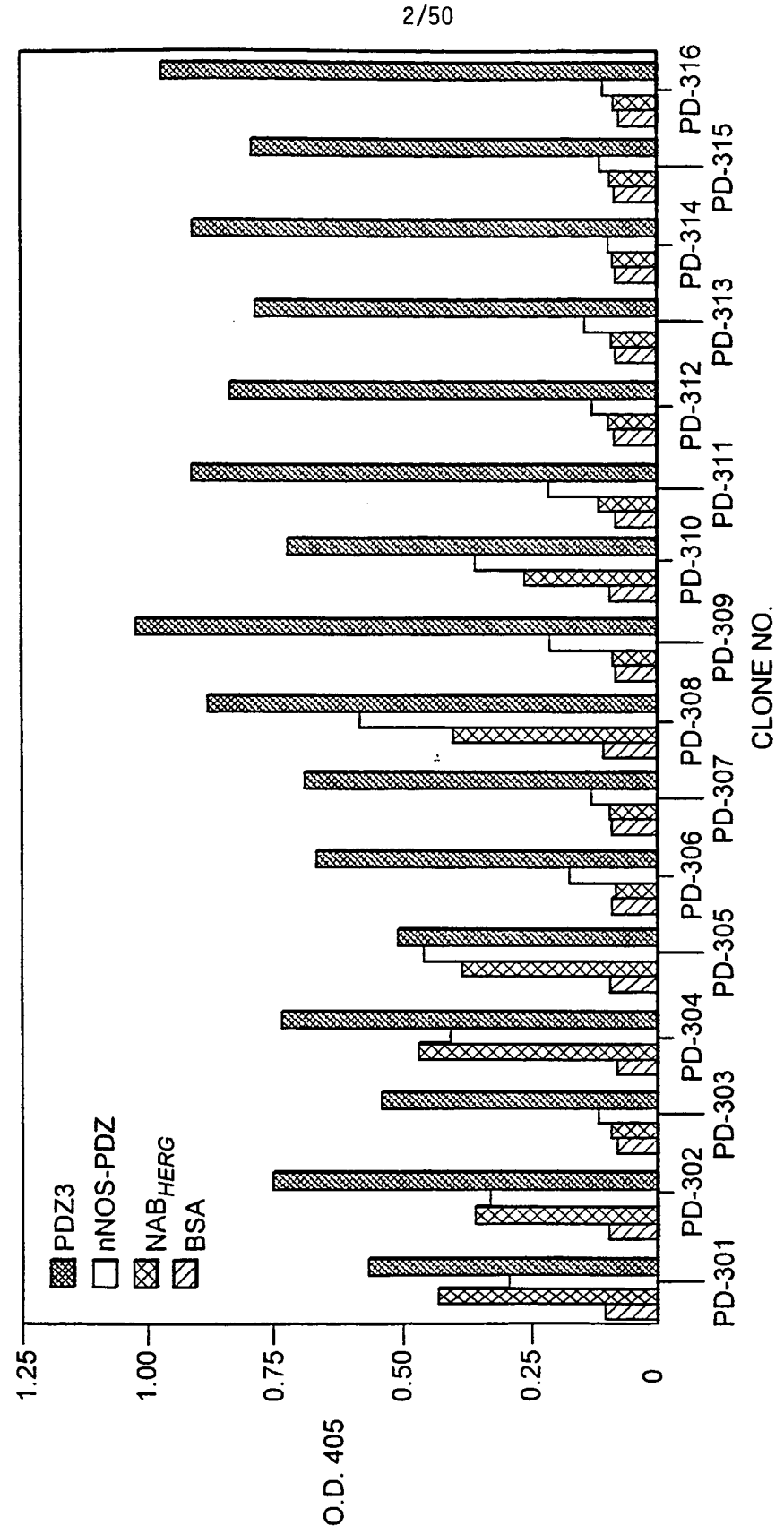
PDZ3
nNOS-PDZ
NAB$_{HERG}$
BSA
O.D. 405
1.25
1.00
0.75
0.50
0.25
0
PD-301
PD-302
PD-303
PD-304
PD-305
PD-306
PD-307
PD-308
PD-309
PD-310
PD-311
PD-312
PD-313
PD-314
PD-315
PD-316
CLONE NO.

FIG. 2A

| CLONE NO. | SEQUENCE |
|---|---|
| Library | GGGXXXXXXXXXXXXXXX* |
| PD-205 | GGGMFVGDQVDLRLETSV* |
| PD-210 | GGGMATSRPSGARRTTSV* |
| PD-211 | GGGMSGWPHDWLGRETTV* |
| PD-212 | GGGMFVGDQVDLRLETSV* |
| PD-215 | GGGILIVRNLETSV* |
| PD-303 | GGGRSLIGAVEKRQETSV* |
| PD-307 | GGGQETLRRLSVGPETSV* |
| PD-312 | GGGHRRSARYLESSV* |
| PD-314 | GGGREASNKVRLRKESTV* |
| PD-315 | GGGGPESLLWKVRRETSL* |
| PD-325 | GGGRIELHGVLKGCETAV* |

FIG. 2B

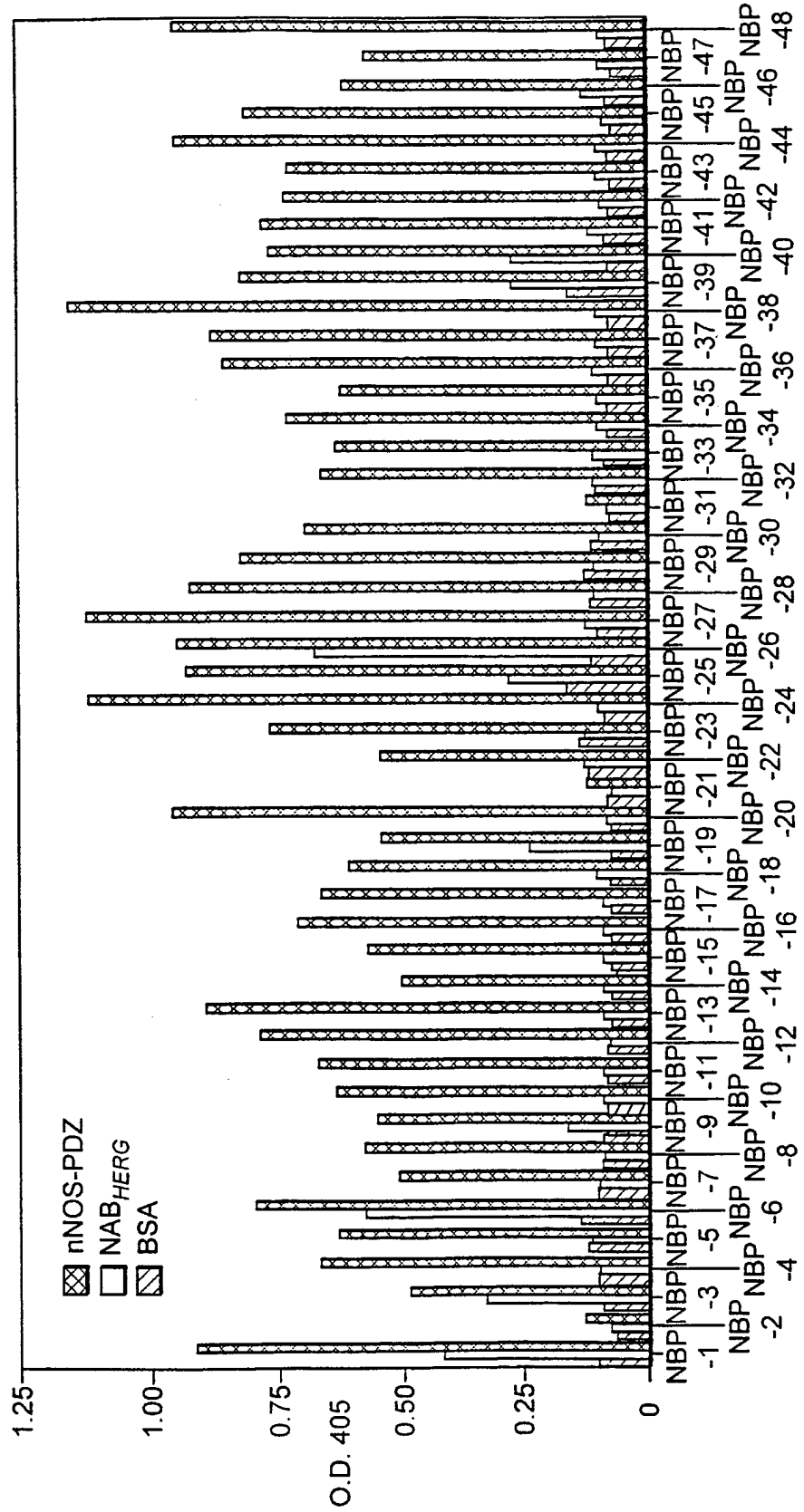
nNOS-PDZ
NAB$_{HERG}$
BSA
O.D. 405
1.25
1.00
0.75
0.50
0.25
0
NBP
-1
-2
-3
-4
-5
-6
-7
-8
-9
-10
-11
-12
-13
-14
-15
-16
-17
-18
-19
-20
-21
-22
-23
-24
-25
-26
-27
-28
-29
-30
-31
-32
-33
-34
-35
-36
-37
-38
-39
-40
-41
-42
-43
-44
-45
-46
-47
-48
CLONE NO.

FIG. 3A

| CLONE NO. | SEQUENCE |
|---|---|
| Library | gggXXXXXXXXXXXXXX* |
| NBP-4 | GGGGTPQKAVHRDWGVSV* |
| NBP-5 | .....GGGIRAGGDPV* |
| NBP-7 | .....GGGDPV* |
| NBP-8 | .GGGDARTKIWNRAADLI* |
| NBP-9 | GGGAQGRWPQFCVYPDAV* |
| NBP-10 | .....GGGVHVFGDSV* |
| NBP-11 | .....GGGVLGDLV* |
| NBP-12 | GGGAMEVTLLSHQPGDPV* |
| NBP-14 | .....GGGDAI* |
| NBP-15 | GGGWAGYGRGMAVSGDMV* |
| NBP-17 | GGGFPFFMGTMGEYGIQV* |
| NBP-18 | GGGLGKDYPSAPDNGDLV* |
| NBP-24 | GGGIYGMMRIGTGLVDVL* |
| NBP-27 | GGGAGQDKQAGQHWGDLV* |
| NBP-28 | .....GGGGVDWV* |
| NBP-32 | .....GGGDAV* |
| NBP-33 | .....GGGRWDWV* |
| NBP-34 | GGGKGHIAITSDGVGDLL* |
| NBP-35 | GGGNYDRVGLLRGPVDFL* |
| NBP-36 | GGGKRPDGVLFQRPGDLV* |
| NBP-37 | .....GGGDAV* |
| NBP-41 | .....GGGDPV* |
| NBP-42 | .....GGGGDAV* |
| NBP-44 | GGGGLARLNLSSYYGDAV* |
| NBP-45 | .....GGGVDWV* |
| NBP-47 | GGGRVIGSPNPSRSADIV* |
| NBP-48 | .....GGGDWV* |
| NBP-49 | GGGSFMNBPVAGTAGDSV* |
| NBP-52 | .....GGGSRGDMV* |
| NBP-53 | .....GGGDWV* |
| NBP-54 | GGGDGMLLRRPQLRWIFC* |
| NBP-55 | .GGGKRDETGFNMWGNAV* |
| NBP-56 | .....GGGWQGDPV* |
| NBP-57 | .....GGGALGDPV* |
| NBP-59 | .....GGGDPV* |
| NBP-60 | .....GGGGDLV* |
| NBP-61 | ..GGGESGSGVRTWGVPV* |
| NBP-62 | ...GGGRVQLVRGGVDCV* |
| NBP-64 | .....GGGDAV* |
| NBP-65 | .GGGWRVVKSVMRWPDPV* |
| NBP-66 | .....GGGDLV* |
| NBP-67 | ..GGGSKSCGRVILGDIV* |
| NBP-68 | .....GGGVDWV* |
| NBP-69 | GGGIIQGQARGTRWGEMV* |
| NBP-70 | .....GGGDAV* |
| NBP-71 | GGGGGWPELNPNLLGVPI* |
| NBP-72 | GGGRCMLNLVTGRWADTV* |
| NBP-73 | GGGGMGQTLEELTTGDWV* |
| NBP-74 | GGGDRGWAVGWGLRGVPV* |
| NBP-76 | .....GGGGPARYGDSV* |
| NBP-77 | .....GGGDLV* |
| NBP-78 | GGGFSSLVLGAGDLGVAP* |
| NBP-79 | .GGGMQWWAQRDLAGDCV* |
| NBP-81 | GGGKDGGRQGANFFGDAV* |
| NBP-82 | .....GGGTWGRAV* |

FIG. 3B

| CLONE NO. | SEQUENCE |
|---|---|
| NBP-83 | GGGLKSTGSEVNSLGDVV* |
| NBP-84 | GGGSEATAVWTSKWSDLV* |
| NBP-85 | GGGPVSSVRYSGVAGDQV* |
| NBP-86 | .....GGGLWSDAV* |
| NBP-87 | GGGRVTGRSSYLGMGDIV* |
| NBP-88 | .....GGGDMV* |
| NBP-89 | GGGKFSVRHTLVSAGDPV* |
| NBP-91 | GGGARGQLPATRCKAFLC* |
| NBP-92 | .....GGGYEEGVAV* |
| NBP-93 | .....GGGDRV* |
| NBP-94 | .....GGGDLV* |
| NBP-95 | GGGVRGALTRGMTPGDPV* |
| NBP-96 | .....GGGDLV* |
| NBP-102 | ....GGGVAGVGKYGDLV* |
| NBP-103 | .....GGGDLV* |
| NBP-107 | .....GGGDVI* |
| NBP-108 | .....GGGKMRVGVDAV* |
| NBP-111 | .....GGGDPV* |
| NBP-112 | ....GGGRDSERLMGIPV* |
| NBP-113 | .....GGGDQV* |
| NBP-114 | .....GGGRWSEGDGV* |
| NBP-117 | GGGLGRGSVRPGRRPDIV* |
| NBP-118 | .....GGGDVV* |
| NBP-119 | GGGIKRLDIYMRNIGDLV* |
| NBP-122 | .....GGGSATAWNGDPV* |
| NBP-123 | ..GGGLDRLRNRVHGDAV* |
| NBP-124 | GGGREVSVCHRPDAGDAV* |
| NBP-125 | GGGSRVPRNTSIFWGNAV* |
| NBP-126 | GGGDCGNVTHAILWGDAV* |
| NBP-128 | GGGKALGAIYVMGGVDAV* |
| NBP-129 | .....GGGWGSPV* |
| NBP-131 | GGGKGSPSLVGPVWADAV* |
| NBP-133 | GGGILNPVPRNLSEGDYV* |
| NBP-136 | .....GGGDQV* |
| NBP-137 | GGGGERLNRSATAGADLV* |
| NBP-138 | .....GGGEGGRNPDIV* |
| NBP-140 | GGGNQRYVVNPFIWGQSV* |
| NBP-142 | ....GGGDSINLSWPVAV* |
| NBP-143 | GGGCMLQVRHIYGPCDAV* |
| NBP-161 | ....GGGVIGKSCYGDAV* |

FIG. 3C

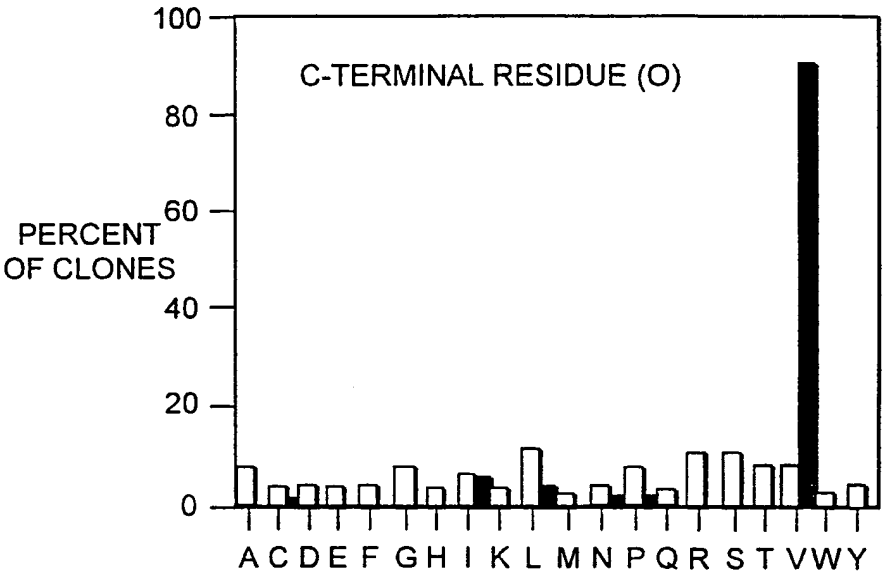
C-TERMINAL RESIDUE (O)
PERCENT OF CLONES
100
80
60
40
20
0
A C D E F G H I K L M N P Q R S T V W Y

FIG. 4A

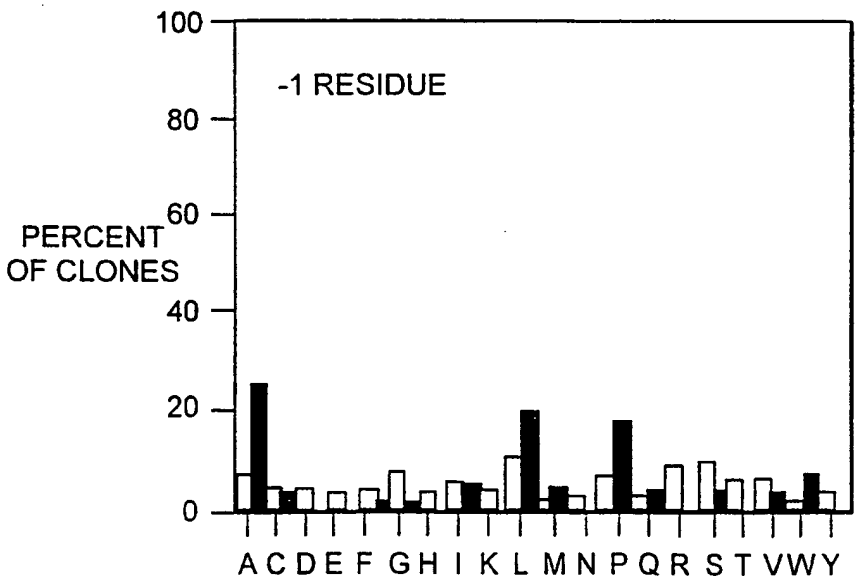
-1 RESIDUE
PERCENT OF CLONES
100
80
60
40
20
0
A C D E F G H I K L M N P Q R S T V W Y

FIG. 4B

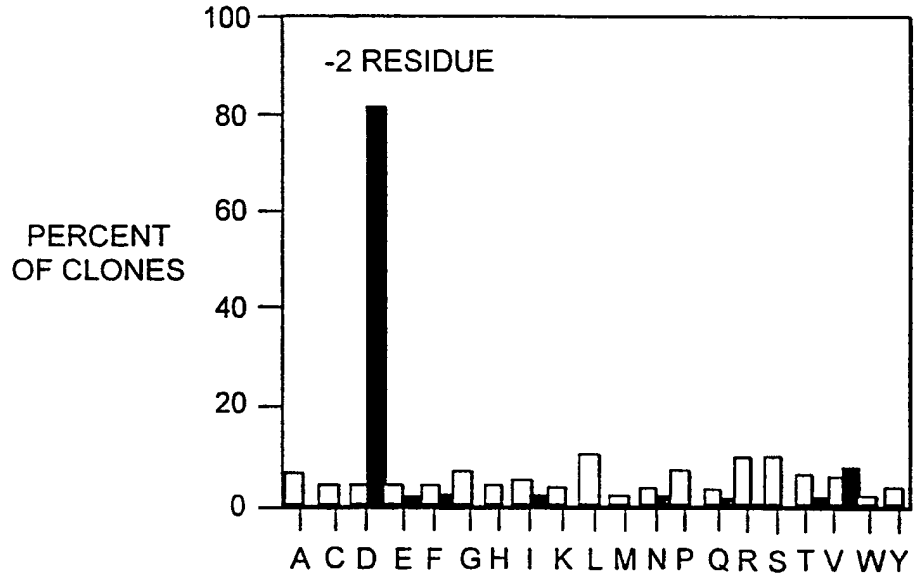
100
80
60
40
20
0
-2 RESIDUE
PERCENT
OF CLONES
A C D E F G H I K L M N P Q R S T V W Y

FIG. 4C

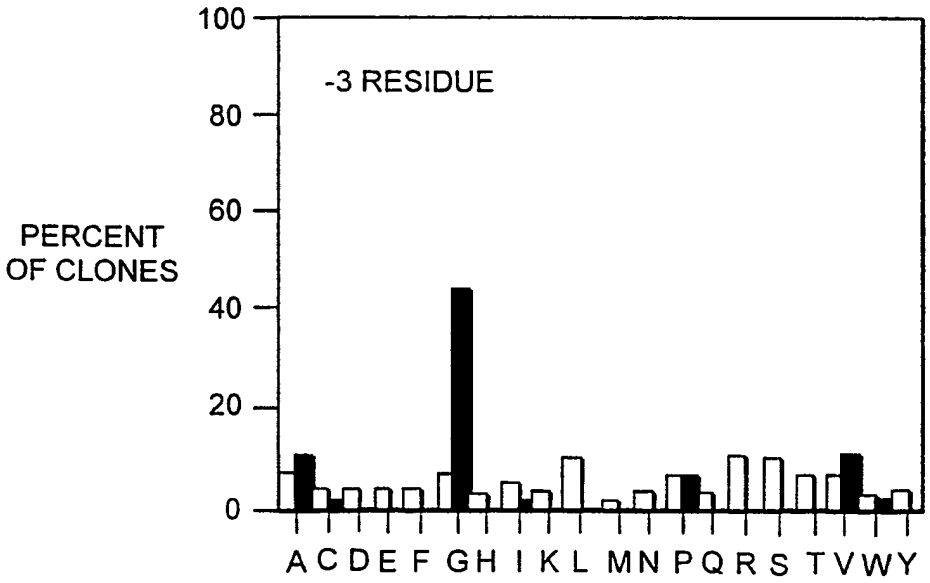
100
80
60
40
20
0
-3 RESIDUE
PERCENT
OF CLONES
A C D E F G H I K L M N P Q R S T V W Y

FIG. 4D

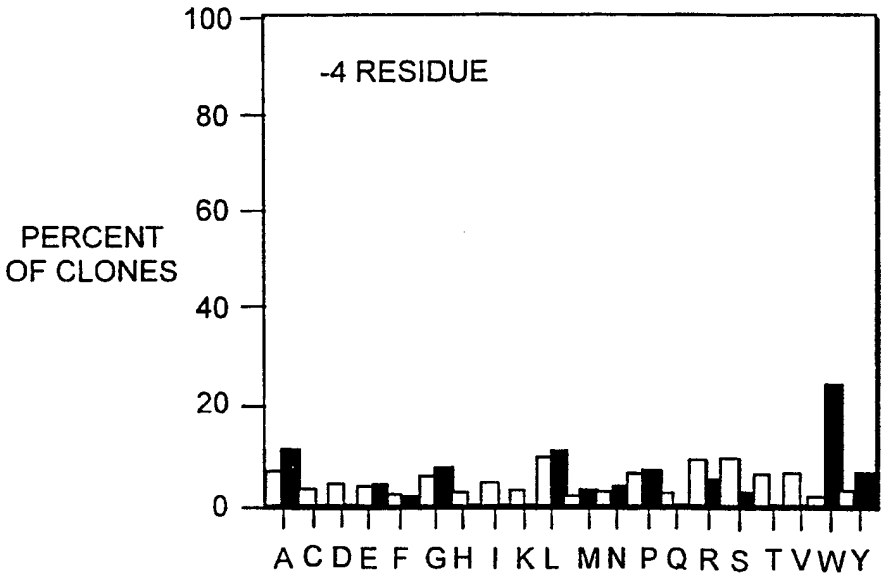
-4 RESIDUE
PERCENT OF CLONES
100
80
60
40
20
0
A C D E F G H I K L MN P Q R S T V W Y

FIG. 4E

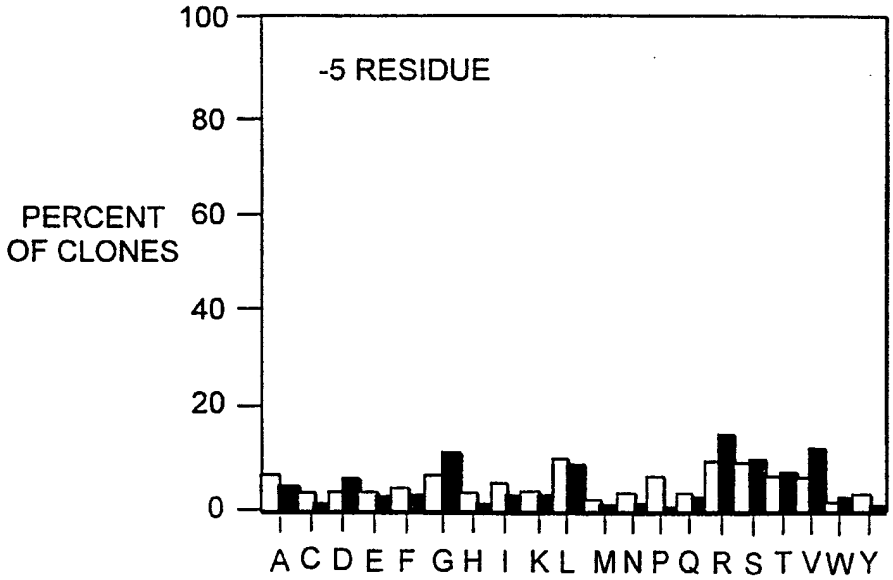
-5 RESIDUE
PERCENT OF CLONES
100
80
60
40
20
0
A C D E F G H I K L MN P Q R S T V W Y

FIG. 4F

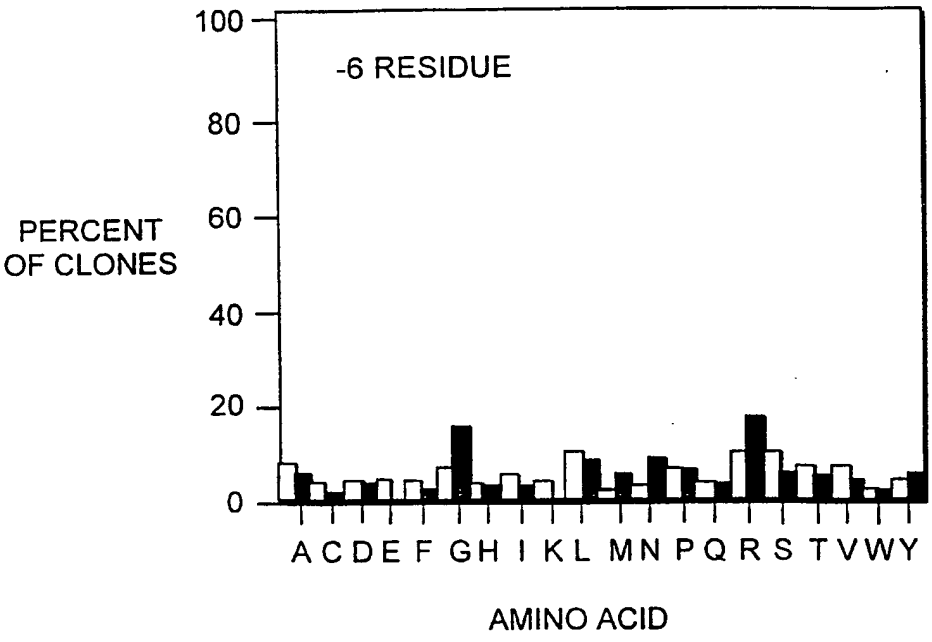
-6 RESIDUE
100
80
60
40
20
0
PERCENT OF CLONES
A C D E F G H I K L M N P Q R S T V W Y
AMINO ACID

FIG. 4G

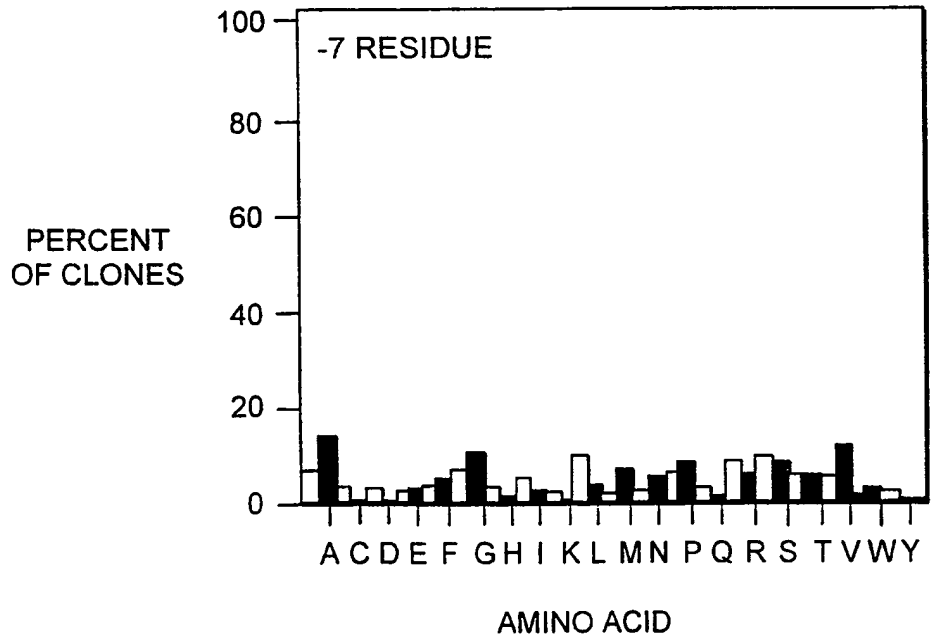
-7 RESIDUE
100
80
60
40
20
0
PERCENT OF CLONES
A C D E F G H I K L M N P Q R S T V W Y
AMINO ACID

FIG. 4H

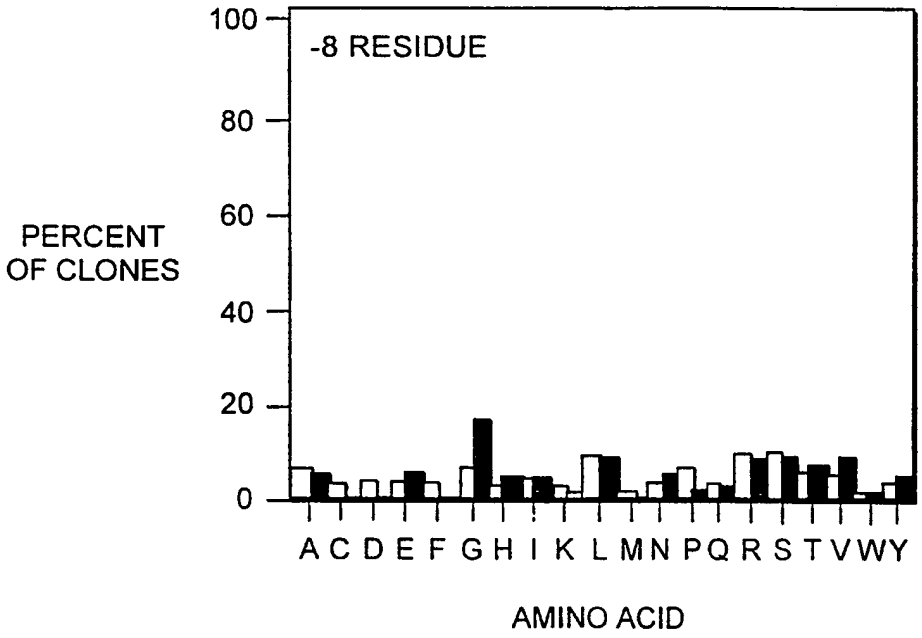
100
80
60
40
20
0
-8 RESIDUE
PERCENT
OF CLONES
A C D E F G H I K L M N P Q R S T V W Y
AMINO ACID

FIG. 4I

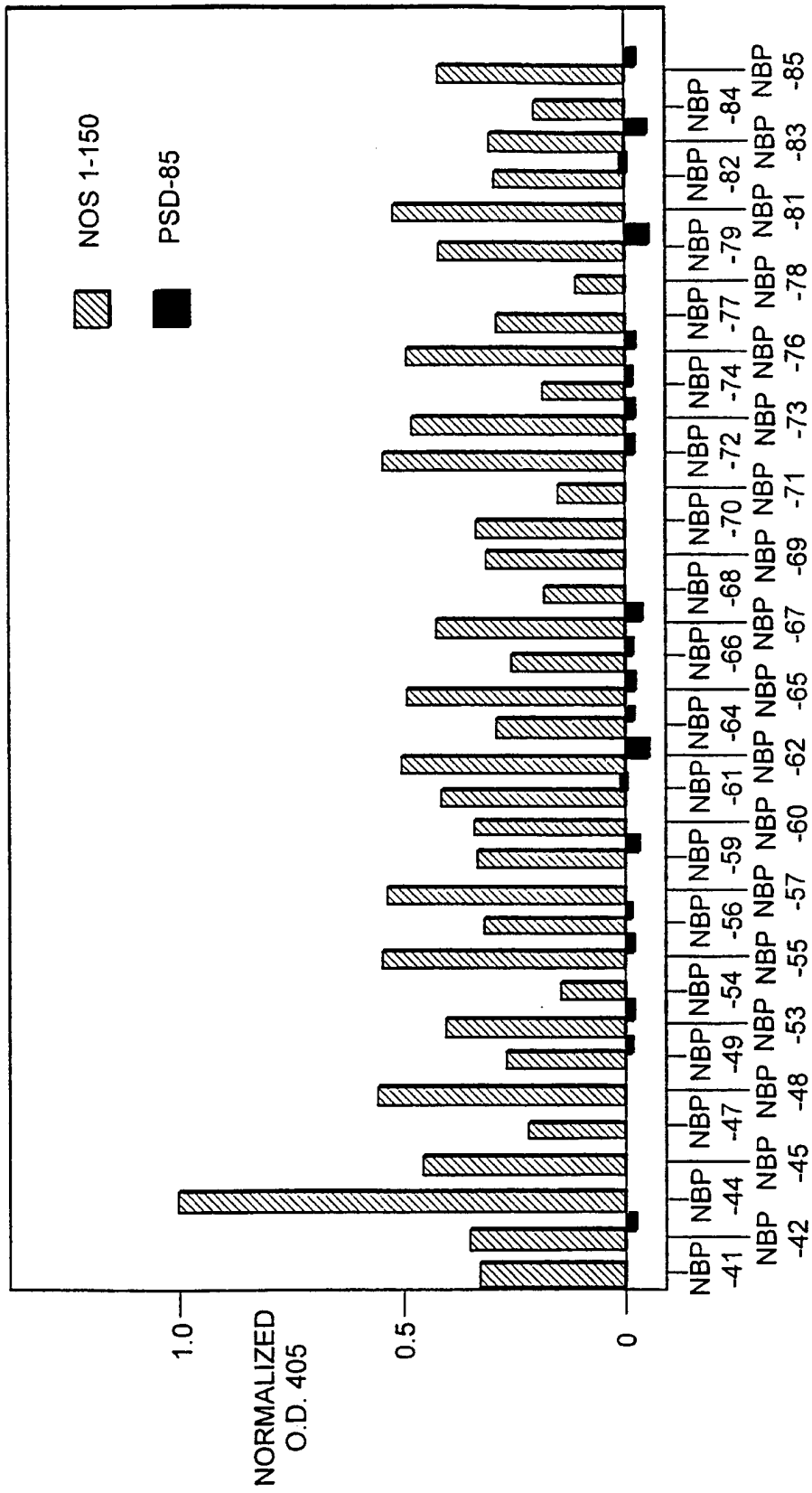
NOS 1-150
PSD-85
NORMALIZED O.D. 405
1.0
0.5
0
NBP -41 NBP -44 NBP -47 NBP -49 NBP -54 NBP -56 NBP -59 NBP -61 NBP -64 NBP -66 NBP -68 NBP -70 NBP -72 NBP -74 NBP -77 NBP -79 NBP -82 NBP -84
NBP -42 NBP -45 NBP -48 NBP -53 NBP -55 NBP -57 NBP -60 NBP -62 NBP -65 NBP -67 NBP -69 NBP -71 NBP -73 NBP -76 NBP -78 NBP -81 NBP -83 NBP -85
CLONE NO.

FIG. 5A

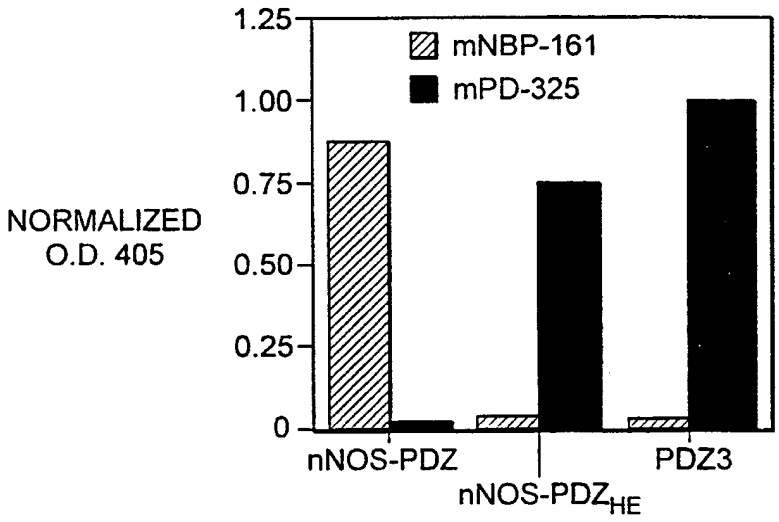
1.25
1.00
0.75
0.50
0.25
0
NORMALIZED
O.D. 405
mNBP-161
mPD-325
nNOS-PDZ
nNOS-PDZHE
PDZ3

FIG. 5B

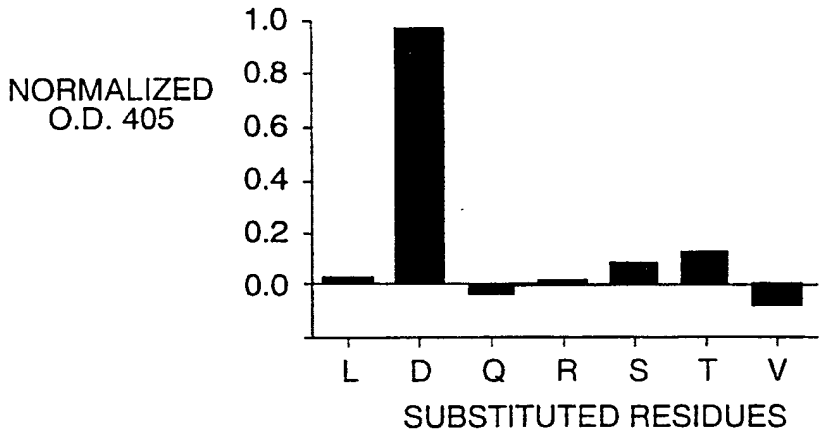
NORMALIZED
O.D. 405
1.0
0.8
0.6
0.4
0.2
0.0
L
D
Q
R
S
T
V
SUBSTITUTED RESIDUES

FIG. 5C

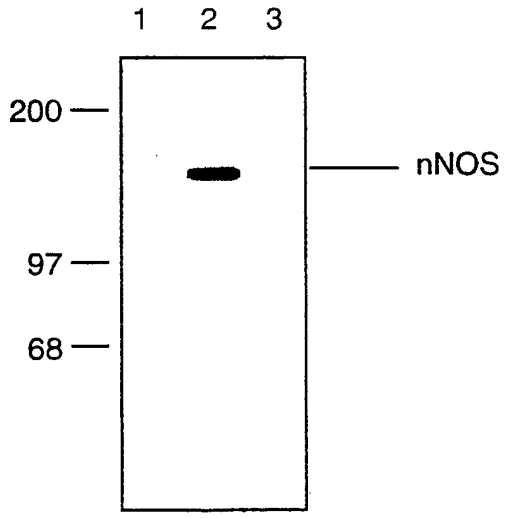
1
2
3
200
97
68
nNOS

FIG. 5D

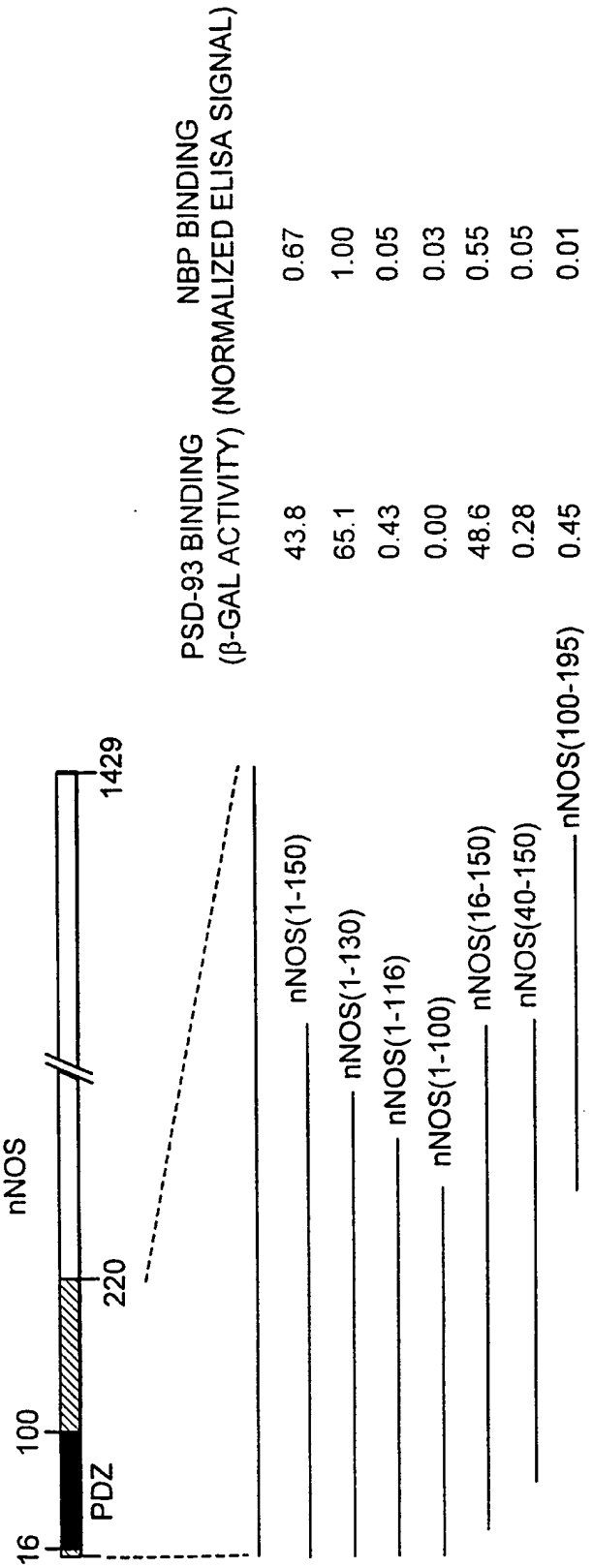
PSD-93 BINDING
(β-GAL ACTIVITY)
NBP BINDING
(NORMALIZED ELISA SIGNAL)
nNOS
16
PDZ
100
220
1429
nNOS(1-150)
43.8
0.67
nNOS(1-130)
65.1
1.00
nNOS(1-116)
0.43
0.05
nNOS(1-100)
0.00
0.03
nNOS(16-150)
48.6
0.55
nNOS(40-150)
0.28
0.05
nNOS(100-195)
0.45
0.01

FIG. 6

1. Endoplasmic reticulum targeting sequence
2. Microbodies C-terminal targeting signal
3. Gram-positive cocci surface proteins anchoring hexapeptide
4. Bipartite nuclear targeting sequence
5. Cell attachment sequence
6. ATP/GTP-binding site motif A (P-loop)
7. Cyclic nucleotide-binding domain signatures
8. EF-hand calcium-binding domain
9. Actinin-type actin-binding domain signatures
10. Anaphylatoxin domain signature and profile
11. Apple domain
12. Band 4.1 family domain signatures
13. C1q domain signature
14. C-terminal cystine knot-signature and profile
15. CUB domain profile
16. Death domain profile
17. EGF-like domain signatures
18. Calcium-binding EGF-like domain signature
19. Forkhead-associated (FHA) domain profile
20. Fibrinogen beta and gamma chains C-terminal domain signature
21. Type II fibronectin collagen-binding domain
22. Hemopexin domain signature
23. Kringle domain signture
24. LDL-receptor class A(LDL RA) domain signature
25. C-type lectin domain signature
26. Osteonectin domain signatures
27. Somatomedin B domain signature
28. Thyroglobulin type-1 repeat signature
29. P-type ("Trefoil") domain signature
30. Cellulose-binding domain, bacterial type
31. Cellulose-binding domain, fungal type
32. Chitin recognition or binding domain signature
33. Barwin domain signatures
34. WAP-type 'four-disulfide core' domain signature
35. Phorbol esters/diacylglycerol binding domain
36. C2 domain signature and profile
37. CAP-Gly domain signature
38. Ly-6/u-PAR domain signature
39. MAM domain signature
40. PH domain profile
41. Phosphotyrosine interaction domain (PLD) profile
42. Src homology 2 (SH2) domain profile
43. Src homology 3 (SH3) domain profile
44. VWFC domain signature
45. WW/rsp5/WWP domain signature and profile
46. ZP domain signature
47. S-layer homology domain signature

FIG. 7

Results: PDZ scan (D-X-V) vs. non-redundant protein database

Human

>gil86707|pir||A31199 125K fusion protein - human gil387034 (M21610) RNA polymerase II [Homo sapiens] (Match DLV)
>gil181745 (M36472) MHC class II cell surface protein [Homo sapiens] (Match DTV)
>gil543732|sp|P36639|8ODP_HUMAN 7,8-DIHYDRO-8-OXOGUANINE TRIPHOSPHATASE (8-OXO-DGTPASE). gil542749|pir||A48886 8-oxo-7,8-dihydroguanosine triphophatase - human gil452589 (D16581) 8-oxo-dGTPase [Homo sapiens] gil1405350 (D38594) 8-oxo-dGTPase [Homo sapiens] (Match DTV)
>gil177776 (M36341) serotonin receptor [Homo sapiens] (Match DGV)
>gil114947|sp|P16278|BGAM_HUMAN BETA GALACTOSIDASE-RELATED PROTEIN PRECURSOR. gil105434|pir||B32633 beta-galactosidase-related protein - human gil179421 (M27508) beta-galactosidase related protein precursor [Homo sapiens] (Match DHV)
>gil187278 (M34150) lysyl oxidase [Homo sapiens] (Match DLV)
>gil553572 (M32887) MHC class II HLA-DQ-alpha-1 [Homo sapiens] (Match DTV)
>gil114940|sp|P16278|BGAL_HUMAN BETA-GALACTOSIDASE PRECURSOR (LACTASE). gil86938|pir||A32611 beta-galactosidase (EC 3.2.1.23) precursor - human gil179401 (M27507) beta-D-galactosidase precursor (EC 3.2.1.23) [Homo sapiens] gil179423 (M34423) beta-galactosidase precursor (EC 3.2.1.23) [Homo sapiens] (Match DHV)
>gil179419 (M22590) beta-galactosidase precursor (EC 3.2.1.23) [Homo sapiens] (Match DHV)
>gil181759 (M63195) DRB1 transplantation antigen [Homo sapiens] (Match DTV)
>gil124462|sp|P17181|INR1_HUMAN INTERFERON-ALPHA/BETA RECEPTOR ALPHA CHAIN PRECURSOR (IFN-ALPHA-REC). gil106790|pir||A32694 interferon alpha receptor precursor - human gil306914 (J03171) interferon-alpha receptor precursor [Homo sapiens] gil156738|gnl|PID|e251628 (A32391) chimeric IFNalpha/beta-receptor [Homo sapiens] (Match DFV)
>gil30972 (Z14206) Ig heavy chain variable region (VDJ) [Homo sapiens] (Match DMV)
>gil32672 (X60459) Human IFNAR gene for interferon alpha/beta receptor [Homo sapiens] (Match DFV)
>gil125472|sp|P10721|KKIT_HUMAN MAST/STEM CELL GROWTH FACTOR RECEPTOR PRECURSOR (SCFR) (PROTO-ONCOGENE TYROSINE-PROTEIN KINASE KIT) (C-KIT) (CD117). gil66311|pir||TVHUKT protein-tyrosine kinase (EC 2.7.1.112) kit precursor - human gil34085 (X06182) protein p145-ckit (AA 1 - 976) [Homo sapiens] gil325686 (X69301) mast/stem cell growth factor receptor [Homo sapiens] (Match DDV)
>gil34992 (X17161) Beta 1-subunit of Na(+),K(+)-ATPase [Homo sapiens] (Match DRV)
>gil631336|pir||S42563 POU domain protein - human gil437809 (Z21963) POU domain protein [Homo sapiens] (Match DVV)
>gil437811 (Z21964) POU domain protein [Homo sapiens] (Match DVV)
>gil437813 (Z21965) POU domain protein [Homo sapiens] (Match DVV)
>gil117098|sp|P20674|COXA_HUMAN CYTOCHROME C OXIDASE POLYPEPTIDE VA PRECURSOR. gil66276|pir||OTHU5A cytochrome-c oxidase (EC 1.9.3.1) chain Va precursor - human gil695360 (M22760) cytochrome c oxidase subunit Va [Homo sapiens] (Match DKV)
**Not human**>gil535709|sp|Q04544|POLN_SOUV3 NON-STRUCTURAL POLYPROTEIN (CONTAINS: RNA-DIRECTED RNA POLYMERASE , THIOL PROTEASE , HELICASE (2C LIKE PROTEIN)). gil476783|pir||A37491 orf1 putative helicase/polymerase polyprotein - Southampton virus gil44364|prf||1906410B rheumatoid factor VH [Homo sapiens] (Match DGV)

Figure 8A

>gil1346544|sp|P48039|ML1A_HUMAN MELATONIN RECEPTOR TYPE 1A (MEL-1A-R). gil602130 (U14108) Mel-1a melatonin receptor [Homo sapiens] (Match DSV)
>gil726255 (U22228) aggrecan [Homo sapiens] (Match DFV)
>gil793763 (D26512) MT-MMP [Homo sapiens] (Match DKV)
>gil804994 (X83535) MT-MMP [Homo sapiens] (Match DKV)
>gil963054 (Z48481) membrane-type matrix metalloproteinase 1 [Homo sapiens]
gil1127837 (U41078) membrane-type matrix metalloproteinase-1 [Homo sapiens] (Match DKV)
>gil976297 (L37839) This CDS feature is included to show the translation of the corresponding V_segment. Presently translation qualifiers on V_segment features are illegal. [Homo sapiens] (Match DAV)
>gil1247461|gnl|PID|e200676 (A26595) interferon beta receptor [Homo sapiens] (Match DFV)
>gil1262584 (D90161) leader sequence, L' [Homo sapiens] (Match DPV)
>gil1495995|gnl|PID|e196537 (X90925) MT-MMP protein [Homo sapiens] (Match DKV)

Mouse

>gil244607|bbs|79586 cleaved prolactin-1, clPRL-1=fragment A [rats, Peptide Partial, 20 aa] (Match DRV)
>gil497021 (U05699) cytochrome c oxidase subunit Va [Mus spretus] (Match DKV)
>gil505029 (D14849) meiosis-specific nuclear structural protein 1 [Mus musculus] (Match DGV)
>gil531881 (U12877) vascular cell adhesion molecule-1 [Mus musculus] (Match DTV)
>gil191913 (M11895) A-1 alpha-amylase [Mus musculus] (Match DKV)
>gil191919 (M11896) B-1 alpha-amylase [Mus musculus] (Match DKV)
>gil192098 (M18187) B144 protein A [Mus musculus] (Match DYV)
>gil196056 (M34984) Ig H-chain [Mus musculus] (Match DTV)
>gil554244 (K03547) myb protein [Mus musculus] (Match DSV)
>gil1363194|pir||A53202 MAMA protein precursor - mouse gil297033 (X67809) mama gene product [Mus musculus] (Match DMV)
>gil423447|pir||S35792 glutamate receptor GluR6C - mouse gil312494 (X66117) glutamate receptor subunit GluR6C [Mus musculus] (Match DTV)
>gil117099|sp|P12787|COXA_MOUSE CYTOCHROME C OXIDASE POLYPEPTIDE VA PRECURSOR. gil90420|pir||S05495 cytochrome-c oxidase (EC 1.9.3.1) chain Va precursor - mouse gil50527 (X15963) cytochrome c oxidase subunit Va preprotein [Mus musculus] (Match DKV)

FIG. 8B-1

>gil805000 (X83536) MT-MMP [Mus musculus] (Match DKV)
>gil939951 (X73037) partial paired box; pid:e74985 [Mus musculus] (Match DGV)
>gil1184877 (U46562) MHC class II transactivator CIITA [Mus musculus] (Match DMV)
>gil1215666 (U17267) T cell receptor-Zeta [Mus musculus] (Match DEV)
>gil1326151 (U52222) Mel-1a melatonin receptor [Mus musculus] (Match DSV)

Rat

>gil666942 (M22615) cholesterol side-chain cleavage enzyme [Rattus norvegicus] (Match DTV)
>gil112437|pir||S20612 triacylglycerol lipase (EC 3.1.1.3) - rat gil56600 (X61925) triacylglycerol lipase [Rattus norvegicus] (Match DTV)
>gil117262|sp|P14137|CPM1_RAT CYTOCHROME P450 XIA1, MITOCHONDRIAL PRECURSOR (P450(SCC)) (CHOLESTEROL SIDE-CHAIN CLEAVAGE ENZYME) (CHOLESTEROL DESMOLASE). gil92074|pir||A34164 cholesterol monooxygenase (side-chain-cleaving) (EC 1.14.15.6) cytochrome P450 11A1 - rat gil203561 (M63133) cytochrome P-450-scc [Rattus norvegicus] gil203639 (J05156) cholesterol side-chain cleavage enzyme precursor (EC 1.14.15.6) [Rattus norvegicus] (Match DTV)
>gil204101 (K01336) beta-fibrinogen [Rattus norvegicus] (Match DKV)
>gil206148 (M16960) calcium-calmodulin-dependent protein kinase II [Rattus norvegicus] (Match DGV)
>gil117100|sp|P11240|COXA_RAT CYTOCHROME C OXIDASE POLYPEPTIDE VA PRECURSOR. gil92182|pir||S04592 cytochrome-c oxidase (EC 1.9.3.1) chain Va precursor - rat gil55971 (X15030) cytochrome c oxidase subunit Va preprotein [Rattus norvegicus] (Match DKV)
>gil682650 (L19118) complement receptor type 1 [Rattus norvegicus] (Match DQV)
>gil805013 (X83537) MT-MMP [Rattus norvegicus] (Match DKV)
>gil1001927 (X91785) membrane-type metalloproteinase [Rattus norvegicus] (Match DKV)

FIG. 8B-2

>gil1334296|gnl||PID|e10391 (X03914) interleukin-3 (aa 102-115) [Rattus norvegicus] (Match DSV)

D. melanogaster

>gil461852|sp|P35220|CTNA_DROME ALPHA-CATENIN. gil422436|pir||A40694 cadherin-associated protein D alpha-catenin - fruit fly (Drosophila melanogaster) gil285752 (D13964) alpha-catenin [Drosophila melanogaster] (Match DAV)
>gil259790|bbs|117942 (S48157) DNA polymerase-primase 180 kda subunit [Drosophila melanogaster, Peptide, 1490 aa] (Match DVV)
>gil546972|bbs|148992 (S70576) putative receptor tyrosine kinase=Dret [Drosophila melanogaster, Canton-S, Peptide Partial, 817 aa] (Match DAV)
>gil321036|pir||PS0443 potassium channel protein Slo G3 - fruit fly (Drosophila melanogaster) (fragment) (Match DLV)

C. elegans

>gil465792|sp|P34428|YL37_CAEEL HYPOTHETICAL 45.5 KD PROTEIN F44B9.7 IN CHROMOSOME III. gil630626|pir||S44810 F44B9.7 protein - Caenorhabditis elegans gil388589 (L23648) putative [Caenorhabditis elegans] (Match DQV)
>gil466054|sp|P34680|YO42_CAEEL HYPOTHETICAL 32.7 KD PROTEIN ZK757.2 IN CHROMOSOME III. gil482218|pir||S41012 hypothetical protein ZK757.2 - Caenorhabditis elegans gil438368 (Z29121) ZK757.2 [Caenorhabditis elegans] (Match DVV)
>gil458953 (U00031) similar to phosphatidylserine decarboxylase [Caenorhabditis elegans] (Match DGV)
>gil722365 (U22833) W02B3.5 [Caenorhabditis elegans] (Match DFV)
>gil746503 (U23516) B0416.2 gene product [Caenorhabditis elegans] (Match DDV)
>gil1019950 (U37429) similar to protein kinase C [Caenorhabditis elegans] (Match DSV)
>gil1055055 (U39850) coded for by C. elegans cDNA yk37g1.5; coded for by C. elegans cDNA yk5c9.5; coded for by C. elegans cDNA yk1a9.5; alternatively spliced form of F52C9.8b [Caenorhabditis elegans] (Match DNV)
>gil1055110 (U39995) coded for by C. elegans cDNA yk25b9.3; coded for by C. elegans cDNA yk25b9.5 [Caenorhabditis elegans] (Match DRV)
>gil1086851 (U41270) Similar to transmembrane domain of family 1 of G-protein coupled receptors. [Caenorhabditis elegans] (Match DEV)
>gil1082139 (Z68118) R01E6.2 [Caenorhabditis elegans] (Match DFV)

FIG. 8C-1

>gil1100868|gnl|PID|e212230 (Z68135) ZK1073.2 [Caenorhabditis elegans] (Match DNV)
>gil1352438|sp|Q10055|IF4N_SCHPO EUKARYOTIC INITIATION FACTOR 4A-LIKE PROTEIN C1F5.10. gil1103737 (Z68136) unknown [Schizosaccharomyces pombe] (Match DMV)
>gil1118060 (U41552) coded for by C. elegans cDNA yk3d11.5; coded for by C. elegans cDNA yk5f4.5 [Caenorhabditis elegans] (Match DIV)
>gil1125770 (U42838) T08G2.2 gene product [Caenorhabditis elegans] (Match DDV)
>gil1185450 (U36581) cyclophilin isoform 9 [Caenorhabditis elegans] (Match DLV)
>gil1229053|gnl|PID|e229193 (Z70207) F15A2.6 [Caenorhabditis elegans] (Match DKV)
>gil1255324 (U51999) C43H6.7 gene product [Caenorhabditis elegans] (Match DIV)
>gil1255397 (U53150) F20A1.2 gene product [Caenorhabditis elegans] (Match DSV)
>gil1313955|gnl|PID|e241752 (Z73098) T21C9.13 [Caenorhabditis elegans] (Match DIV)
>gil1627717|gnl|PID|e276022 (Z81053) E02A10.4 [Caenorhabditis elegans] (Match DIV)
>gil1627903|gnl|PID|e275743 (Z81076) F35C5.f [Caenorhabditis elegans] (Match DGV)
>gil1658357 (U64849) K04A8.8 gene product [Caenorhabditis elegans] (Match DKV)

S. cerevisiae

>gil728821|sp|P39010|AKR1_YEAST ANKYRIN REPEAT-CONTAINING PROTEIN AKR1. gil626094|pir||S48521 AKR1 protein - yeast (Saccharomyces cerevisiae) gil466522 (L31407) ankyrin repeat-containing protein [Saccharomyces cerevisiae] gil1230637 (U51030) Ankyrin repeat-

FIG. 8C-2

containing protein (Swiss Prot. accession number P39010). [Saccharomyces cerevisiae] gi|1586336|prf||2203403A ankyrin repeat-containing protein [Saccharomyces cerevisiae] (Match DMV)
>gi|731840|sp|P40500|YII9_YEAST HYPOTHETICAL 23.9 KD PROTEIN IN SGA1-THS1 INTERGENIC REGION. gi|1077785|pir||S49791 hypothetical protein YI9910.07 - yeast (Saccharomyces cerevisiae) gi|577125 (Z46728) YI9910.07, unknown orf, len: 205, CAI: 0.11 [Saccharomyces cerevisiae] gi|763257 (Z47047) unknown [Saccharomyces cerevisiae] (Match DEV)
>gi|140345|sp|P25554|YCB0_YEAST HYPOTHETICAL 16.6 KD PROTEIN IN GBP2-PEL1 INTERGENIC REGION. gi|83138|pir||S19337 hypothetical protein YCL010c - yeast (Saccharomyces cerevisiae) gi|5358|gnl|PID|e264452 (X59720) YCL010c, len:146 [Saccharomyces cerevisiae] (Match DTV)
>gi|731426|sp|P39941|YEI0_YEAST HYPOTHETICAL 56.5 KD PROTEIN IN HXT8 5'REGION. gi|1077619|pir||S50519 hypothetical protein YEL070w - yeast (Saccharomyces cerevisiae) gi|603248 (U18795) Yel070p [Saccharomyces cerevisiae] gi|1302610|gnl|PID|e239852 (Z71688) ORF YNR073c [Saccharomyces cerevisiae] (Match DQV)
>gi|1174566|sp|P41896|T2FB_YEAST TRANSCRIPTION INITIATION FACTOR IIF, BETA SUBUNIT (TFIIF-BETA) (TFIIF MEDIUM SUBUNIT) (TRANSCRIPTION FACTOR G 54 KD SUBUNIT). gi|1078424|pir||B55482 transcription initiation factor IIF 54K chain - yeast (Saccharomyces cerevisiae) gi|639703 (U13016) transcription initiation factor TFIIF middle subunit [Saccharomyces cerevisiae] (Match DVV)
>gi|825501 (L42348) HOL1 [Saccharomyces cerevisiae] (Match DGV)
>gi|258767|bbs|117066 cytochrome c oxidase VIa subunit homolog [Saccharomyces cerevisiae, JHRY1-2 alpha, Peptide Partial, 19 aa, segment 1 of 5] (Match DKV)
>gi|847740 (U19781) beta-fructofuranosidase 2 precursor [Saccharomyces cerevisiae] (Match DTV)
>gi|914979 (U32445) P8283.8 gene product [Saccharomyces cerevisiae] (Match DRV)
>gi|1353041|sp|P46984|YJS4_YEAST HYPOTHETICAL 13.6 KD PROTEIN IN SWE1-ATP12 INTERGENIC REGION. gi|1077849|pir||S56967 hypothetical protein YJL184w - yeast (Saccharomyces cerevisiae) gi|1008389 (Z49459) ORF YJL184w; pid:e201216 [Saccharomyces cerevisiae] (Match DAV)
>gi|1352875|sp|P47104|YJ03_YEAST HYPOTHETICAL 154.9 KD PROTEIN IN MER2-PET191 INTERGENIC REGION. gi|1077878|pir||S57052 hypothetical protein YJR033c - yeast (Saccharomyces cerevisiae) gi|1015679 (Z49533) ORF YJR033c; pid:e203690 [Saccharomyces cerevisiae] (Match DFV)
>gi|1129167 (X87297) J1590 gene product [Saccharomyces cerevisiae] (Match DFV)
>gi|1134890 (Z68290) Akr1p [Saccharomyces cerevisiae] gi|1226040 (Z70202) Akr1p [Saccharomyces cerevisiae] (Match DMV)

FIG. 8D-1

>gi|1302574|gnl|PID|e239841 (Z71670) ORF YNR055c [Saccharomyces cerevisiae] (Match DGV)
>gi|1322879|gnl|PID|e243887 (Z72748) ORF YGL226w [Saccharomyces cerevisiae] (Match DLV)
>gi|1322961|gnl|PID|e243366 (Z72790) ORF YGR005c [Saccharomyces cerevisiae] (Match DVV)
>gi|1323286|gnl|PID|e243550 (Z72948) ORF YGR163w [Saccharomyces cerevisiae] (Match DDV)
>gi|1420794|gnl|PID|e252191 (Z75275) ORF YOR367w [Saccharomyces cerevisiae] (Match DIV)

Other

>gi|401194|sp|P31015|TNA2_SYMTH TRYPTOPHANASE 2 (L-TRYPTOPHAN INDOLE-LYASE 2). gi|477858|pir||B49022 tryptophanase (EC 4.1.99.1) Tna2 - Symbiobacterium thermophilum gi|216979 (D10013) tryptophanase [Symbiobacterium thermophilum] (Match DLV)
>gi|155612 (L09651) phosphoglycerate mutase [Zymomonas mobilis] (Match DLV)
>gi|1361344|pir||D36891 transfer complex protein TrsC - Staphylococcus aureus gi|310610 (L11998) putative [Staphylococcus aureus] gi|405562 (L19570) putative [Plasmid pSK41] gi|739958|prf||2004267D membrane protein traC [Staphylococcus sp.] (Match DDV)
>gi|625710|pir||C49695 4-methyl-5-(beta-hydroxyethyl)thiazole monophosphate synthesis protein ThiF - Escherichia coli gi|414234 (M88701) thiF [Escherichia coli] (Match DPV)
>gi|97777|pir||A38729 pyruvate decarboxylase (EC 4.1.1.1) - Sarcina ventriculi (fragment) gi|249565|bbs|103674 pyruvate decarboxylase {EC 4.1.1.1} [Sarcina ventriculi, strain JK, Peptide Partial, 36 aa] (Match DYV)

FIG. 8D-2

>gi|298240|bbs|125733 DNA polymerase homolog [bacterium-like organism, citrus greening disease-associated, Peptide, 207 aa] (Match DLV)
>gi|477173|pir||A48368 N5,N10-methenyltetrahydromethanopterin cyclohydrolase - Archaeoglobus fulgidus (fragment) gi|299881|bbs|130469 N5,N10-methenyltetrahydromethanopterin cyclohydrolase {N-terminal} [Archaeoglobus fulgidus, VC-19, DSM 4304, Peptide Partial, 38 aa] (Match DGV)
>gi|406020 (U01764) unknown [Mycoplasma genitalium] (Match DSV)
>gi|414513 (U02113) homology to ribosomal protein L1 Z11839 [Mycoplasma genitalium] (Match DVV)
>gi|396331 (U00006) similar to E. coli ChlN [Escherichia coli] (Match DPV)
>gi|543897|sp|P35804|BLIP_STRCL BETA-LACTAMASE INHIBITORY PROTEIN PRECURSOR (BLIP). gi|98890|pir||A36710 beta-Lactamase inhibitory protein precursor - Streptomyces clavuligerus gi|153192 (M34538) beta-lactamase inhibitory protein precursor [Streptomyces clavuligerus] (Match DLV)
>gi|538757|pir||A53488 heat shock cognate protein 66 - Escherichia coli gi|454766 (U05338) Hsc66 [Escherichia coli] (Match DEV)
>gi|461079|bbs|142342 GroEL homolog {N-terminal} [Francisella tularensis, LVS, Peptide Partial, 18 aa] (Match DGV)
>gi|547685|sp|P36541|HSCA_ECOLI HEAT SHOCK PROTEIN HSCA (HSC66). gi|1073308|pir||B36958 66K hsp70 homolog HscA - Escherichia coli gi|402675 (U01827) Hsp70 [Escherichia coli] (Match DEV)
>gi|129002|sp|P07061|NYLB_FLASP 6-AMINOHEXANOATE-DIMER HYDROLASE (NYLON OLIGOMERS DEGRADING ENZYME EII). gi|77553|pir||A29516 6-aminohexanoate-dimer hydrolase (EC 3.5.1.46) EII - Flavobacterium sp. KI72 plasmid pOAD2 gi|43418 (X00046) EII enzyme [Flavobacterium sp.] gi|488340 (D26094) 6-aminohexanoate-dimer hydrolase [Flavobacterium sp.] gi|223803|prf||0912258A enzyme RSIIA,nylon degrading [Flavobacterium sp.] (Match DAV)
>gi|488342 (D26094) 6-aminohexanoate-dimer hydrolase [Flavobacterium sp.] (Match DAV)
>gi|507769 (U09675) RNA polymerase beta subunit [Liberobacter africanum] (Match DGV)
>gi|118911|sp|P10740|DPSD_ECOLI PHOSPHATIDYLSERINE DECARBOXYLASE PROENZYME. gi|78759|pir||A29234 phosphatidylserine decarboxylase (EC 4.1.1.65) precursor - Escherichia coli gi|537004 (U14003) phosphatidylserine decarboxylase [Escherichia coli] gi|551827 (J03916) phosphatidylserine decarboxylase [Escherichia coli] (Match DQV)
>gi|1361237|pir||S56466 phosphotransferase system trehalose permease - Escherichia coli gi|537082 (U14003) phosphotransferase system trehalose permease [Escherichia coli] (Match DIV)
>gi|479220|pir||S32798 merR protein - Xanthomonas sp. transposon Tn5053 gi|480554|pir||S37035 regulatory protein merR - Alcaligenes sp.

FIG. 8E-1

gi|480563|pir||S37044 regulatory protein merR - Pseudomonas fluorescens gi|1086170|pir||S51756 regulatory protein merR - Pseudomonas testosteroni gi|154910 (L03729) putative [Transposon Tn5053] gi|388554 (L20693) mer operon regulator [Alcaligenes sp.] gi|393198 (L20694) mer operon regulator [Plasmid pMER05] gi|397588 (Z23094) merR regulatory protein (repressor /inducer) [Alcaligenes sp.] gi|397618 (Z23095) merR regulatory protein (repressor /inducer) [Pseudomonas fluorescens] gi|483767 (X73112) mercury resistance DNA-binding protein [Pseudomonas fluorescens] gi|607170 (Z33481) regulatory protein [Comamonas testosteroni] gi|710575 (L40585) merR regulatory protein (repressor /inducer) [Transposon Tn5053] (Match DAV)
>gi|142082 (L02520) ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit [Anabaena sp.] gi|142086 (L02521) ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit [Anabaena sp.] gi|142088 (L02522) ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit [Anabaena sp.] gi|142105 (J01540) ribulose-1,5-bisphosphate carboxylase large subunit (rbcL) [Anabaena sp.] (Match DTV)
>gi|1075610|pir||S52644 phycobilisome maturation protein - Synechococcus sp. gi|142130 (M94218) phycobilisome maturation protein [Anacystis nidulans] gi|446765|prf||1912291J phycobilisome maturation protein [Synechococcus sp.] (Match DRV)
>gi|466182|sp|P35151|YPUB_BACSU HYPOTHETICAL 7.2 KD PROTEIN IN PPIB-SIPS INTERGENIC REGION (ORFX1). gi|629118|pir||S45538 hypothetical protein X1 - Bacillus subtilis gi|410120 (L09228) ORFX1 [Bacillus subtilis] (Match DRV)
>gi|142967 (M17642) succinate dehydrogenase [Bacillus subtilis] (Match DRV)
>gi|118613|sp|P08066|DHSB_BACSU SUCCINATE DEHYDROGENASE IRON-SULFUR PROTEIN. gi|1075923|pir||B27763 succinate dehydrogenase (EC 1.3.99.1) iron-sulfur protein - Bacillus subtilis gi|143527 (M13470) iron-sulfur protein [Bacillus subtilis] (Match DRV)
>gi|144453 (M94320) very similar to DNA polymerase of Bacillus subtilis bacteriophage SPO2; potential DNA polymerase; putative [Citrus greening disease-associated bacterium-like organism] (Match DLV)

FIG. 8E-2

>gi|78587|pir||G25035 hypothetical protein 2 - Escherichia coli plasmid ColIa gi|455439 (M13819) ORF2 [Plasmid ColIa] (Match DDV)
>gi|78588|pir||H25035 hypothetical protein 2 - Escherichia coli plasmid ColIb gi|455441 (M13820) ORF2 [Plasmid ColIb] (Match DDV)
>gi|145313 (K01304) L-ribulokinase (araB) [Escherichia coli] (Match DSV)
>gi|120350|sp|P26608|FLIS_ECOLI FLAGELLAR PROTEIN FLIS. gi|145989 (M85240) flagellar protein [Escherichia coli] (Match DPV)
>gi|125924|sp|P26593|LACD_LACLA TAGATOSE 1,6-DIPHOSPHATE ALDOLASE. gi|97943|pir||D39778 LacD tagatose-1,6-diphosphate aldolase - Lactococcus lactis gi|149396 (M65190) lacD [Lactococcus lactis] gi|149409 (M60447) tagatose 1,6-diP aldolase [Lactococcus lactis] (Match DKV)
>gi|68525|pir||SYEXI isoleucine--tRNA ligase (EC 6.1.1.5) - Methanobacterium thermoautotrophicum gi|149728 (M59245) transfer RNA-Ile synthetase [Methanobacterium thermoautotrophicum] (Match DKV)
>gi|150352 (M84113) ORF1 [Transposon mini-Tn3Cm] (Match DAV)
>gi|121875|sp|P24375|GVPK_HALHA GVPK PROTEIN. gi|81055|pir||JQ1128 GvpK protein - Halobacterium halobium plasmid pNRC100 gi|43524 (X55648) gvpK gene product [Halobacterium halobium] gi|455299 (M58557) gas vesicle protein [Plasmid pNRC100] (Match DDV)
>gi|127013|sp|P13111|MERR_SERMA MERCURIC RESISTANCE OPERON REGULATORY PROTEIN. gi|96175|pir||A33858 merR protein - Escherichia coli plasmid pDU1358 gi|455313 (M24940) mercury resistance protein [Plasmid pDU1358] (Match DAV)
>gi|150838 (K02336) EII enzyme (6-aminohexanoic acid linear oligomer hydrolase) [Plasmid pOAD2] (Match DAV)
>gi|294462 (M28607) insB [Escherichia coli] (Match DKV)
>gi|121389|sp|P13556|GLNB_RHOCA NITROGEN REGULATORY PROTEIN P-II. gi|151934 (M28244) glutamine synthetase glnB (EC 6.3.1.2) [Rhodobacter capsulatus] gi|829596 (U25953) PII protein [Rhodobacter capsulatus] (Match DAV)
>gi|135828|sp|P27477|THTR_SYNP7 PUTATIVE THIOSULFATE SULFURTRANSFERASE PRECURSOR (RHODANESE-LIKE PROTEIN). gi|280211|pir||A43669 rhodanese homolog rhdA precursor - Synechococcus sp. gi|154604 (M65244) rhdA [Synechococcus sp.] (Match DRV)
>gi|731176|sp|P40981|XYLR_THER8 PUTATIVE XYLOSE REPRESSOR. gi|632297|pir||S41787 xylR protein - Thermophilic bacterium gi|311188 (L18965) putative xylose repressor gene; putative [Thermophilic bacterial sp.] (Match DYV)
>gi|1175762|sp|P46015|YDEB_ANASP HYPOTHETICAL PROTEIN IN DEVB 5'REGION. gi|556606 (U14553) ORF [Anabaena sp.] (Match DYV)
>gi|1072948|pir||S51047 mauR protein - Paracoccus denitrificans gi|558803 (U12464) LysR-type transcriptional activator [Paracoccus denitrificans] (Match DAV)

FIG. 8F-1

>gi|629404|pir||S48833 cytochrome-c3 hydrogenase (EC 1.12.2.1) alpha chain - Pyrococcus furiosus gi|563905 (X75255) hydrogenase (alpha subunit) [Pyrococcus furiosus] (Match DGV)
>gi|130794|sp|P07781|PQQ2_ACICA COENZYME PQQ SYNTHESIS PROTEIN II. gi|95318|pir||E32252 gene II protein - Acinetobacter calcoaceticus gi|38744 (X06452) gene II [Acinetobacter calcoaceticus] (Match DLV)
>gi|128258|sp|P10996|NIFE_CLOPA NITROGENASE IRON-MOLYBDENUM COFACTOR BIOSYNTHESIS PROTEIN NIFE. gi|80505|pir||S04079 nitrogenase (EC 1.18.6.1) molybdenum-iron protein nifE - Clostridium pasteurianum gi|40587 (X13606) NifE protein (AA 1 - 456) [Clostridium pasteurianum] (Match DYV)
>gi|547614|sp|P36553|HEM6_ECOLI COPROPORPHYRINOGEN III OXIDASE, AEROBIC (COPROPORPHYRINOGENASE) (COPROGEN OXIDASE). gi|1073344|pir||B36964 coproporphyrinogen oxidase (EC 1.3.3.3), aerobic - Escherichia coli gi|453969 (X75413) coproporphyrinogen oxidase [Escherichia coli] (Match DWV)
>gi|95681|pir||S06878 beta-Galactosidase (EC 3.2.1.23) - Escherichia coli (fragment) gi|41904 (X16313) lacZ 5'-region [Escherichia coli] (Match DGV)
>gi|78569|pir||S04774 hypothetical protein - Escherichia coli (fragment) gi|42746 (X15859) open reading frame (122 AA); pid:g42746 [Escherichia coli] (Match DQV)
>gi|129003|sp|P07062|NYLC_FLASP 6-AMINOHEXANOATE-DIMER HYDROLASE (NYLON OLIGOMERS DEGRADING ENZYME EII'). gi|77554|pir||B22644 6-aminohexanoate-dimer hydrolase (EC 3.5.1.46) EII' - Flavobacterium sp. plasmid pOAD2 gi|43420 (X02864) EII' (aa 1-392)

FIG. 8F-2

[Flavobacterium sp.] gi|223804|prf||0912258B enzyme RSIIB,nylon degrading [Flavobacterium sp.] (Match DAV)
>gi|79956|pir||JH0207 hypothetical 10.8K protein - Enterococcus faecalis plasmid pAM-beta-1 gi|45739 (X17092) ORFF (ttg start codon) [Enterococcus faecalis] (Match DFV)
>gi|114867|sp|P26177|BCHX_RHOCA CHLOROPHYLLIDE REDUCTASE 35.5 KD CHAIN (CHLORIN REDUCTASE). gi|79513|pir||S17823 protochlorophyllide reductase (EC 1.3.1.33) 35.5K chain - Rhodobacter capsulatus gi|46131 (Z11165) 333 aa (35.5 kD) chlorophillide reductase subunit, also known as chlorophyll Fe protein [Rhodobacter capsulatus] (Match DDV)
>gi|116927|sp|P24716|COPR_STRAG PLASMID COPY CONTROL PROTEIN COPR. gi|98007|pir||S22829 hypothetical protein - Streptococcus agalactiae gi|581557 (X62150) 92 aa polypeptide [Streptococcus agalactiae] gi|769739 (X72021) circular [Streptococcus agalactiae] (Match DFV)
>gi|134993|sp|P09398|STRG_STRGR STREPTOMYCIN BIOSYNTHESIS PROTEIN STRG. gi|80801|pir||S17777 strG protein - Streptomyces griseus gi|49266 (Y00459) strG [Streptomyces griseus] (Match DTV)
>gi|401018|sp|P31814|RPOB_THECE DNA-DIRECTED RNA POLYMERASE SUBUNIT B. gi|280354|pir||S25563 DNA-directed RNA polymerase (EC 2.7.7.6) chain B - Thermococcus celer gi|48140 (X67313) Subunit B of DNA-dependent RNA polymerase [Thermococcus celer] (Match DRV)
>gi|625666|pir||A36925 LysR-type transcriptional activator CbbR - Xanthobacter flavus gi|581832 (Z22705) DNA-binding protein [Xanthobacter flavus] (Match DPV)
>gi|515608 (Z35397) C. sativus 3-ketoacyl-CoA thiolase [Arabidopsis thaliana] (Match DIV)
>gi|451328 (U02021) ecdysteroid receptor [Aedes aegypti] (Match DQV)
>gi|413919 (D21101) Guanyl Cyclase [Hemicentrotus pulcherrimus] (Match DDV)
>gi|514269 (U07706) dihydropteroate synthetase [Plasmodium falciparum] (Match DQV)
>gi|505159 (Z30659) dihydropteroate synthetase [Plasmodium falciparum]
gi|505169 (Z30665) dihydropteroate synthetase [Plasmodium falciparum]
gi|505171 (Z30655) dihydropteroate synthetase [Plasmodium falciparum]
gi|505175 (Z30657) dihydropteroate synthetase [Plasmodium falciparum] (Match DQV)
>gi|505161 (Z30660) dihydropteroate synthetase [Plasmodium falciparum] (Match DQV)
>gi|505163 (Z30653) dihydropteroate synthetase [Plasmodium falciparum] (Match DQV)

FIG. 8G-1

>gil505165 (Z30664) dihydropteroate synthetase [Plasmodium falciparum] (Match DQV)
>gil630466|pir||S47154 dihydropterin pyrophosphokinase/dihydropteroate synthetase - Plasmodium falciparum gil505179 (Z31584) Dihydropterin pyrophosphokinase and Dihydropteroate synthetase [Plasmodium falciparum] (Match DQV)
>gil505167 (Z30654) dihydropteroate synthetase [Plasmodium falciparum] gil505173 (Z30656) dihydropteroate synthetase [Plasmodium falciparum] (Match DQV)
>gil505177 (Z30658) dihydropteroate synthetase [Plasmodium falciparum] (Match DQV)
>gil585279|sp|Q08169|HUGA_APIME HYALURONOGLUCOSAMINIDASE PRECURSOR (HYALURONIDASE) (ALLERGEN API M II) (API M 2). gil476996|pir||A47477 hyaluronidase - honeybee gil155680 (L10710) hyaluronidase [Apis mellifera] (Match DQV)
>gil159276 (M64611) putative [Hydra vulgaris] (Match DVV)
>gil552162 (L28823) reverse transcriptase [Phlebotomus perniciosus] (Match DTV)
>gil160301 (M15212) glycophorin binding protein [Plasmodium falciparum] (Match DEV)
>gil118063|sp|P16065|CYGS_STRPU SPERACT RECEPTOR PRECURSOR (GUANYLATE CYCLASE). gil279588|pir||OYURCP speract receptor precursor - sea urchin (Strongylocentrotus purpuratus) gil161477 (M22444) guanylate cyclase [Strongylocentrotus purpuratus] (Match DDV)
>gil556182 (L36665) ORF; putative [Gonyaulax polyedra] (Match DLV)
>gil163188 (L06320) alpha-interferon receptor [Bos taurus] (Match DSV)
>gil246581|bbs|86109 zona pellucida-binding protein, AWN-1=C13' fragment [swine, sperm, Peptide Partial, 10 aa] (Match DXV)
>gil399217|sp|P30932|CD9_BOVIN CD9 ANTIGEN. gil89462|pir||JX0221 CD9 antigen - bovine gil162821 (M81720) CD9 antigen [Bos taurus] (Match DMV)
>gil562100 (U15975) putative brain ryanodine receptor [Sus scrofa] (Match DQV)
>gil462415|sp|Q04790|INR1_BOVIN INTERFERON-ALPHA/BETA RECEPTOR ALPHA CHAIN PRECURSOR (IFN-ALPHA-REC). gil346520|pir||S27387 interferon alpha receptor type 1 - bovine gil432 (X68443) interferon receptor type 1 [Bos taurus] (Match DSV)

FIG. 8G-2

>gi|137049|sp|P01145|UR1_CATCO UROTENSIN I. gi|69066|pir||UOCC1M urotensin I - white sucker gi|268092|gb|I02277| Sequence 2 from Patent US 4528189 gi|270944|gb|I01722| Sequence 2 from Patent US 4908352 (Match DEV)
>gi|268113|gb|I02366| Sequence 1 from Patent US 4533654 gi|268114|gb|I02367| Sequence 2 from Patent US 4533654 (Match DEV)
>gi|268397|gb|I03062| Sequence 4 from Patent US 4605642 (Match DEV)
>gi|268996|gb|I00642| Sequence 8 from Patent US 4742157 (Match DXV)
>gi|270945|gb|I01724| Sequence 3 from Patent US 4908352 (Match DEV)
>gi|227991|prf||1714327A urotensin I [Hippoglossoides elassodon]
gi|270946|gb|I01726| Sequence 4 from Patent US 4908352 (Match DEV)
>gi|592318|gb|I11163| Sequence 4 from Patent WO 8906658 (Match DSV)
>gi|593118|gb|I10347| Sequence 3 from Patent WO 8705938 (Match DTV)
>gi|594746|gb|I04467| Sequence 7 from Patent EP 0162738 (Match DGV)
>gi|132051|sp|P00875|RBL_SPIOL RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN PRECURSOR. gi|68133|pir||RKSPL ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain precursor - spinach chloroplast gi|231312|pdb|8RUB|L Ribulose 1,5-Bisphosphate Carboxylase(Slash)oxygenase (E.C.4.1.1.39) Complex With Co2,Mg++ And 2-Carboxyarabinitol-1,5-Bisphosphate gi|12291 (V00168) ribulose 1,5-bisphophate carboxylase [Spinacia oleracea] gi|343375 (J01443) ribulose bisphosphate carboxylase large subunit [Spinacia oleracea] (Match DTV)
>gi|111564|pir||S09074 cytochrome P450-4b - rat (fragment) (Match DGV)
>gi|82261|pir||S06161 chitinase (EC 3.2.1.14) - potato (fragment) gi|21465 (X14133) endochintinase (315 AA) [Solanum tuberosum] (Match DTV)
>gi|84502|pir||B28563 hemoglobin chain IV - earthworm (Lumbricus terrestris) (fragment) (Match DDV)
>gi|84636|pir||S00492 hemocyanin chain Ia - Japanese spiny lobster (fragment) (Match DDV)
>gi|320206|pir||S28389 acyl carrier protein - Escherichia coli (fragment) (Match DTV)
>gi|281333|pir||PQ0397 nonstructural protein NS5 - hepatitis C virus (isolate E-b12) (fragment) (Match DPV)
>gi|538860|pir||A61213 photoreaction center protein H - Rhodospirillum rubrum gi|227675|prf||1709158B puh gene [Rhodospirillum rubrum] (Match DRV)
>gi|97994|pir||G35905 hypothetical protein 1 (Sm2) - Streptococcus mutans (Match DIV)
>gi|79995|pir||A28551 hypothetical protein 1 - Streptococcus mutans (strain GS-5) gi|1196925 (M18954) unknown protein [Streptococcus mutans] (Match DIV)
>gi|483018|pir||B47607 immunogenic protein MPB70/MPB80 - Mycobacterium bovis (strain BCG) (fragment) (Match DPV)

FIG. 8H-1

>gi|1174853|sp|P42743|UBCY_ARATH UBIQUITIN-CONJUGATING ENZYME E2-18 KD (UBIQUITIN-PROTEIN LIGASE) (UBIQUITIN CARRIER PROTEIN) (PM42). gi|481811|pir||S39483 ubiquitin-conjugating enzyme UBC2-1 - Arabidopsis thaliana gi|22658 (X68306) ubiquitin-conjugating enzyme [Arabidopsis thaliana] (Match DKV)
>gi|225491|prf||1304301B glycoprotein S8 [Brassica rapa] (Match DLV)
>gi|137055|sp||UR1_PLAFE_2 [Segment 2 of 2] UROTENSIN I PRECURSOR. gi|280657|pir||A43978 urotensin I - European flounder gi|227317|prf||1701464A urotensin I [Platichthys flesus] (Match DEV)
>gi|87715|pir||PH0159 HLA-DRB sigma antigen DRB1-0701-Dw17 - human (Match DTV)
>gi|87718|pir||PT0162 HLA-DRB sigma antigen DRB1-0901-Dw23 - human (Match DTV)
>gi|91588|pir||PT0641 T-cell receptor beta chain V-D-J region (120-2R) - mouse (fragment) (Match DWV)
>gi|481922|pir||S40164 hemagglutinin-neuraminidase - Newcastle disease virus gi|437889 (X71994) hemagglutinin-neuraminidase [Newcastle disease virus] (Match DGV)
>gi|1169937|sp|P43519|GLNB_RHOSH NITROGEN REGULATORY PROTEIN P-II (PII SIGNAL TRANSDUCING PROTEIN). gi|421339|pir||S33180 glnB protein - Rhodobacter sphaeroides gi|809751 (X71659) glnB gene product [Rhodobacter sphaeroides] gi|1586928|prf||2205239A Glu synthetase [Rhodobacter sphaeroides] (Match DAV)
>gi|98843|pir||S14091 40K protein - Saccharopolyspora erythraea (Match DAV)
>gi|479179|pir||S32438 pol polyprotein - Volvox carteri retrotransposon VCRT-I-1 (fragment) gi|938289 (X69621) reverse transcriptase [Volvox carteri] (Match DDV)
>gi|479181|pir||S32440 pol polyprotein - Volvox carteri retrotransposon VCRT-I-3 (fragment) gi|938291 (X69623) reverse transcriptase [Volvox carteri] (Match DDV)
>gi|479183|pir||S32442 pol polyprotein - Volvox carteri retrotransposon VCRT-I-6 (fragment) gi|938294 (X69626) reverse transcriptase [Volvox carteri] (Match DDV)

FIG. 8H-2

>gi|479184|pir||S32443 pol polyprotein - Volvox carteri retrotransposon VCRT-I-8 (fragment) (Match DDV)
>gi|479185|pir||S32444 pol polyprotein - Volvox carteri retrotransposon VCRT-II-1 (fragment) gi|938295 (X69629) reverse transcriptase [Volvox carteri] (Match DDV)
>gi|479188|pir||S32447 pol polyprotein - Volvox carteri retrotransposon VCRT-II-4 (fragment) gi|938298 (X69632) reverse transcriptase [Volvox carteri] (Match DDV)
>gi|479190|pir||S32449 pol polyprotein - Volvox carteri retrotransposon VCRT-II-3 (fragment) gi|938297 (X69631) reverse transcriptase [Volvox carteri] (Match DDV)
>gi|477748|pir||B47759 reverse transcriptase (copia-like retrotransposon) - upland cotton (fragment) gi|167317 (M94472) reverse transcriptase [Gossypium hirsutum] (Match DDV)
>gi|1076316|pir||S51478 Di19 protein - Arabidopsis thaliana gi|469110 (X78584) Di19 [Arabidopsis thaliana] (Match DEV)
>gi|99777|pir||S14951 S-locus-specific glycoprotein SLG-8 - field mustard gi|17708 (X55274) S-locus glycoprotein [Brassica campestris] (Match DLV)
>gi|478421|pir||JQ2380 S-locus-specific glycoprotein precursor - rape gi|1076455|pir||S42280 S-locus glycoprotein - rape gi|167170 (L08608) S-locus glycoprotein [Brassica napus] gi|904227 (L10736) S-locus related glycoprotein [Brassica napus] (Match DLV)
>gi|99826|pir||S24546 S-locus glycoprotein - rape gi|17868 (Z11725) S-locus glycoprotein [Brassica napus] (Match DLV)
>gi|434858 (X76472) pid:g434858 [Crucianella angustifolia] (Match DAV)
>gi|478565|pir||S10849 alpha-amylase/trypsin inhibitor - durum wheat (Match DYV)
>gi|1172751|sp|P41390|PUR1_SCHPO AMIDOPHOSPHORIBOSYLTRANSFERASE (GLUTAMINE PHOSPHORIBOSYLPYROPHOSPHATE AMIDOTRANSFERASE) (ATASE). gi|481335|pir||S38482 amidophosphoribosyltransferase (EC 2.4.2.14) - fission yeast (Schizosaccharomyces pombe) gi|629904|pir||S43526 PRPP amidotransferase (EC 2.4.2.14) - yeast (Schizosaccharomyces pombe) gi|410512 (X72293) PRPP amidotransferase [Schizosaccharomyces pombe] (Match DFV)
>gi|542640|pir||A48810 fibrinogen B beta subunit - African clawed frog (fragment) gi|450951 (U05035) fibrinogen B-beta subunit [Xenopus laevis] (Match DDV)
>gi|477549|pir||A49192 transthyretin - bullfrog (fragment) gi|299846|bbs|130235 transthyretin, T-T3BP=3,5,3'-L-triiodothyronine-specific binding protein {N-terminal} [bullfrogs, tadpole plasma, Peptide Partial, 19 aa] (Match DAV)
>gi|481489|pir||S38695 class II histocompatibility antigen betea chain - slender loris (fragment) (Match DTV)

FIG. 8I-1

>gi|478168|pir||E49164 chromogranin-B - rat (fragment) gi|239365|bbs|66367 chromogranin-B, CgB=glucagonoma peptide [rats, Peptide Partial, 38 aa] (Match DNV)
>gi|543521|pir||B61222 cytochrome-c oxidase (EC 1.9.3.1) chain II - mitochondrion Steinernema intermedii (SGC4) (fragment) (Match DEV)
>gi|543672|pir||JQ2350 protein kinase (EC 2.7.1.37) - turkey herpesvirus gi|406788 (X68653) protein kinase homologue [Gallid herpesvirus 2] gi|583811 (A18267) ORF5 [Gallid herpesvirus 2] gi|1253294|pat|US|5470734|5 Sequence 5 from patent US 5470734 (Match DSV)
>gi|478188|pir||F47758 reverse transcriptase (copia-like retrotransposon) - Liriodendron chinense (fragment) gi|168306 (M94477) reverse transcriptase [Liriodendron chinense] (Match DDV)
>gi|116359|sp|P23472|CHLY_HEVBR HEVAMINE A (CHITINASE / LYSOZYME. gi|82026|pir||S17205 chitinase (EC 3.2.1.14) hevamine - Para rubber tree gi|234388|bbs|52808 hevamine [Hevea brasiliensis, Peptide Partial, 273 aa] gi|1311006|pdb|1HVQ| Glycosidase, Chitin Degradation, Multifunctional Enzyme Mol_id: 1; Molecule: Hevamine A; Chain: Null; Ec: 3.2.1.14, 3.2.1.17; Heterogen: N-,N'-,N''-Triacetyl-Chitotriose; Other_details: Plant EndochitinaseLYSOZYME gi|1311007|pdb|1HVM| Glycosidase, Chitin Degradation, Multifunctional Enzyme Mol_id: 1; Molecule: Hevamine A; Chain: Null; Ec: 3.2.1.14, 3.2.1.17; Other_details: Plant EndochitinaseLYSOZYME gi|1421554|pdb|1LLO| Hevamine A (A Plant EndochitinaseLYSOZYME) COMPLEXED WITH Allosamidin Chitinase, Lysozyme Mol_id: 1; Molecule: Hevamine; Chain: Null; Synonym: ChitinaseLYSOZYME; Ec: Ec 3.2.1.14, 3.2.1.17; Heterogen: Allosamidin (Match DSV)
>gi|467822 (U02606) chitinase [Solanum tuberosum] (Match DTV)
>gi|629717|pir||S43317 chitinase (EC 3.2.1.14) - potato (fragment) gi|467824 (U02607) chitinase [Solanum tuberosum] (Match DTV)
>gi|467911 (U03086) ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit [Sarcothalia decipiens] (Match DVV)
>gi|514215 (U02963) dynein beta heavy chain [Chlamydomonas reinhardtii] (Match DVV)

FIG. 8I-2

>gil516552 (U10078) cyclin IbZm [Zea mays] (Match DLV)
>gil1170247|sp|P43082|HEVL_ARATH HEVEIN-LIKE PROTEIN PRECURSOR.
gil407248 (U01880) pre-hevein-like protein [Arabidopsis thaliana] (Match DRV)
>gil625982|pir||JC2250 S-locus-specific glycoprotein S12 precursor - field mustard
gil547238|bbs|149323 (S70937) S-glycoprotein [Brassica campestris, S12S12 homozygotes, stigmas, Peptide, 436 aa] gil743639|prf||2013216A S glycoprotein [Brassica rapa] (Match DLV)
>gil289868 (L12640) ribulose 1,5-bisphosphate carboxylase large subunit [Chloranthus japonicus] (Match DTV)
>gil460648 (L29492) ribulose 1,5 bisphosphate carboxylase [Comesperma ericinum] (Match DTV)
>gil290939 (L12649) ribulose 1,5-bisphosphate carboxylase large subunit [Hedyosmum arborescens] (Match DTV)
>gil310368 (L19972) ribulose 1,5-bisphosphate carboxylase [Stegolepis allenii] (Match DKV)
>gil484236 (L05041) ribulose 1,5-bisphosphate carboxylase large subunit [Tradescantia sp.] (Match DKV)
>gil166459 (L06946) beta-tubulin [Acremonium uncinatum] (Match DAV)
>gil166467 (L06954) beta-tubulin [Acremonium sp.] gil168130 (L06959) beta-tubulin [Epichloe amarillans] (Match DAV)
>gil119975|sp|P16972|FER_ARATH FERREDOXIN PRECURSOR.
gil99692|pir||S09979 ferredoxin [2Fe-2S] precursor - Arabidopsis thaliana gil16437 (X51370) ferredoxin precursor [Arabidopsis thaliana] gil166698 (M35868) ferrodoxin A [Arabidopsis thaliana] (Match DIV)
>gil167172 (M36301) S-6-glycoprotein [Brassica campestris]
gil225490|prf||1304301A glycoprotein S6 [Brassica rapa] (Match DLV)
>gil166461 (L06951) beta-tubulin [Acremonium coenophialum] gil166463 (L06952) beta-tubulin [Acremonium sp.] gil166469 (L06963) beta-tubulin [Acremonium sp.] gil166471 (L06964) beta-tubulin [Acremonium coenophialum] gil168122 (L06955) beta-tubulin [Epichloe festucae] gil168124 (L06956) beta-tubulin [Epichloe festucae] gil168126 (L06957) beta-tubulin [Epichloe festucae] gil168128 (L06958) beta-tubulin [Epichloe amarillans] gil168133 (L06961) beta-tubulin [Epichloe amarillans] gil168135 (L06962) beta-tubulin [Epichloe sp.] (Match DAV)
>gil169359 (J01262) phaseolin [Phaseolus vulgaris] gil897800 (V01163) phaseolin [Phaseolus vulgaris] (Match DDV)
>gil457400 (D21840) MAP kinase [Arabidopsis thaliana] (Match DSV)
>gil310372 (L13485) ribulosebisphosphate carboxylase [Sphagnum palustre] (Match DTV)
>gil309636 (L11058) 'Ribulosebiphosphate Carboxylase' [Ophioglossum engelmanii] (Match DTV)
>gil3815 (X00788) 1G2 protein [Schizophyllum sp.] (Match DPV)

FIG. 8J-1

>gi|547991|sp|P36606|NAH_SCHPO PROBABLE NA(+)/H(+) ANTIPORTER. gi|82816|pir||S20951 Na+/H+ antiporter - fission yeast (Schizosaccharomyces pombe) gi|5090 (Z11736) putative sodium/proton antiporter [Schizosaccharomyces pombe] (Match DYV)
>gi|134531|sp|P22553|SLS2_BRAOA S-LOCUS-SPECIFIC GLYCOPROTEIN BS29-2 PRECURSOR. gi|17889 (X16123) S locus specific glycoprotein [Brassica oleracea] (Match DLV)
>gi|17894 (X55275) S-locus glycoprotein [Brassica oleracea] (Match DLV)
>gi|134534|sp|P07761|SLS6_BRAOL S-LOCUS-SPECIFIC GLYCOPROTEIN S6 PRECURSOR (SLSG-6). gi|81703|pir||A27827 S-locus-specific glycoprotein S6 precursor - wild cabbage gi|17901 (Y00268) SLSG (AA -31 to 405) [Brassica oleracea] gi|225542|prf||1305350A protein,S locus allele [Brassica oleracea var. botrytis] (Match DLV)
>gi|436130 (X76634) ribulose-1,5-bisphosphate carboxylase [Physcomitrella patens] (Match DTV)
>gi|1346963|sp|P20455|RBL_ATRRS RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN PRECURSOR. gi|99516|pir||F34921 ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - Atriplex rosea chloroplast gi|11323 (X55831) rubisco large subunit [Atriplex rosea] (Match DTV)
>gi|131998|sp|P19163|RBL_NEUMU RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN PRECURSOR. gi|68147|pir||RKNULM ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain precursor - Neurachne munroi chloroplast gi|100640|pir||H34921 ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - Neurachne munroi chloroplast gi|11751 (X55828) rubisco large subunit [Neurachne munroi] (Match DKV)
>gi|131999|sp|P19164|RBL_NEUTE RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN PRECURSOR. gi|68146|pir||RKNULT ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain precursor - Neurachne tenuifolia chloroplast gi|100641|pir||G34921 ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - Neurachne tenuifolia chloroplast gi|11798 (X55827) rubisco large subunit [Neurachne tennifolia] (Match DKV)

FIG. 8J-2

>gi|299258|bbs|127093 (S56181) pyruvate dehydrogenase alpha subunit {C-terminal} {EC 1.2.4.1} [human, Peptide Partial Mutant, 14 aa] (Match DQV)
>gi|385595|bbs|133340 (S62078) platelet-derived growth factor A-chain, PDGF A-chain {N-terminal} [human, Peptide Partial, 53 aa] (Match DSV)
>gi|124884|sp|P16808|IR10_HCMVA HYPOTHETICAL PROTEIN IRL10 PRECURSOR (TRL10). gi|76487|pir||S09903 hypothetical protein IRL10 precursor - human cytomegalovirus (strain AD169) gi|833108 (X17403) HCMVIRL10 = TRL10 (AA 1-171) [Human cytomegalovirus] (Match DNV)
>gi|134532|sp|P17840|SLS3_BRAOL S-LOCUS-SPECIFIC GLYCOPROTEIN S13 PRECURSOR (SLSG-13). gi|81698|pir||B27827 S-locus-specific glycoprotein S13 precursor - wild cabbage (fragment) (Match DLV)
>gi|134533|sp|P17841|SLS4_BRAOL S-LOCUS-SPECIFIC GLYCOPROTEIN S14 PRECURSOR (SLSG-14). gi|81699|pir||C27827 S-locus-specific glycoprotein S14 precursor - wild cabbage (fragment) (Match DIV)
>gi|267240|sp|P30088|UPA2_HUMAN UNKNOWN PROTEIN FROM 2D-PAGE OF PLASMA (SPOT 10). (Match DQV)
>gi|117097|sp|P00426|COXA_BOVIN CYTOCHROME C OXIDASE POLYPEPTIDE VA. gi|66277|pir||CABO cytochrome-c oxidase (EC 1.9.3.1) chain Va - bovine gi|229632|prf||771727A oxidase heme a,cytochrome [Bos taurus] (Match DKV)
>gi|126902|sp|P80040|MDH_CHLAU MALATE DEHYDROGENASE. (Match DIV)
>gi|126903|sp|P80039|MDH_CHLTE MALATE DEHYDROGENASE. (Match DVV)
>gi|126906|sp|P80037|MDH_HELGE MALATE DEHYDROGENASE. (Match DIV)
>gi|131906|sp|P00879|RBL_ANASP RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN PRECURSOR. gi|68158|pir||RKAIL7 ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - Anabaena sp. gi|223640|prf||0904327A carboxylase,RBP [Anabaena sp.] (Match DTV)
>gi|417995|sp|P30138|THIF_ECOLI THIF PROTEIN. (Match DPV)
>gi|136991|sp|P16787|UL96_HCMVA HYPOTHETICAL PROTEIN UL96. gi|76602|pir||S09861 hypothetical protein UL96 - human cytomegalovirus (strain AD169) gi|833080 (X17403) HCMVUL96 (AA 1-115) [Human cytomegalovirus] (Match DAV)
>gi|137504|sp|P21075|VB17_VACCC PROTEIN B17. gi|93309|pir||G42527 B17L protein - vaccinia virus (strain Copenhagen) gi|335564 (M35027) B17L; putative [Vaccinia virus] (Match DNV)
>gi|267281|sp|Q01221|VB17_VACCV PROTEIN B17. gi|321391|pir||JQ1810 B16L protein - vaccinia virus (strain WR) gi|222761 (D11079) 39.5K protein [Vaccinia virus] (Match DNV)

FIG. 8K-1

>gi|138210|sp|P18538|VGLB_HSVMD GLYCOPROTEIN B PRECURSOR. gi|73946|pir||VGBERB glycoprotein B precursor - Marek's disease virus (strain RB1B) gi|221837 (D13713) glycoprotein B precursor [Gallid herpesvirus type 1] gi|1100890 (U39846) glycoprotein B [Gallid herpesvirus 2] (Match DAV)
>gi|547619|sp|P12554|HEMA_NDVA HEMAGGLUTININ-NEURAMINIDASE. gi|67467|pir||HNNZAV hemagglutinin-neuraminidase (EC 3.2.1.-) - Newcastle disease virus (strain Australia-Victoria virulent) (Match DGV)
>gi|547620|sp|P35740|HEMA_NDVC HEMAGGLUTININ-NEURAMINIDASE. gi|419457|pir||C36829 hemagglutinin-neuraminidase (EC 3.2.1.-) - Newcastle disease virus (strain CHI/85) gi|332352 (M24716) hemagglutinin-neuraminidase [Newcastle disease virus] (Match DRV)
>gi|547621|sp|P35741|HEMA_NDVH3 HEMAGGLUTININ-NEURAMINIDASE. gi|419459|pir||A36829 hemagglutinin-neuraminidase (EC 3.2.1.-) - Newcastle disease virus (strain HER/33) (Match DGV)
>gi|122996|sp|P12556|HEMA_NDVI HEMAGGLUTININ-NEURAMINIDASE. gi|77139|pir||S07126 hemagglutinin-neuraminidase (EC 3.2.1.-) - Newcastle disease virus (strain Italien) gi|332362 (M18640) hemagglutinin-neuraminidase [Newcastle disease virus] gi|226158|prf||1413194A hemagglutinin neuraminidase [Newcastle disease virus] (Match DGV)
>gi|547622|sp|P35742|HEMA_NDVJ HEMAGGLUTININ-NEURAMINIDASE. gi|419460|pir||D36829 hemagglutinin-neuraminidase (EC 3.2.1.-) - Newcastle disease virus (strain IBA/85) gi|332354 (M24717) hemagglutinin-neuraminidase [Newcastle disease virus] (Match DRV)
>gi|122997|sp|P12557|HEMA_NDVM HEMAGGLUTININ-NEURAMINIDASE. gi|332368 (M19479) hemagglutinin-neuraminidase glycoprotein [Newcastle disease virus] (Match DKV)
>gi|135128|sp|P26499|SYI_METTH ISOLEUCYL-TRNA SYNTHETASE (ISOLEUCINE--TRNA LIGASE) (ILERS). (Match DKV)
>gi|401222|sp|P31779|TTHY_RANCA TRANSTHYRETIN (PREALBUMIN) (TADPOLE T3-BINDING PROTEIN) (T-T3BP). (Match DAV)

FIG. 8K-2

>gi|465058|sp|P33878|VB17_VARV PROTEIN B17. gi|419242|pir||I36856 B18L protein - variola virus (strain India-1967) gi|628217|pir||S46875 gene B17L protein (COP) - variola virus gi|439093 (L22579) homolog of vaccinia virus CDS B17L; putative [variola major virus] gi|457077 (X69198) pid:g457077 [Variola virus] gi|516436 (X67117) B17L COP gene product [Variola virus] gi|885783 (U18339) D6L [Variola virus] gi|885845 (U18341) B15L [Variola virus] gi|1150675 (X72086) ORF17L; B18L in citation [3] [Variola virus] gi|745309|prf||2015436HK B18L gene [Variola major virus] (Match DNV)
>gi|138975|sp|P15775|VSMP_CVBF PUTATIVE SMALL MEMBRANE PROTEIN (NONSTRUCTURAL PROTEIN NS3) (NONSTRUCTURAL 9.7 KD PROTEIN). gi|74875|pir||MNIHB3 nonstructural protein NS3 - bovine coronavirus (strain F15) gi|58686 (X51347) NS3 protein (AA 1-84) [Bovine coronavirus] (Match DDV)
>gi|138976|sp|P15779|VSMP_CVBM PUTATIVE SMALL MEMBRANE PROTEIN (NONSTRUCTURAL PROTEIN NS3) (NONSTRUCTURAL 9.7 KD PROTEIN). gi|418984|pir||D46346 nonstructural protein NS3 - bovine coronavirus (strain Mebus) gi|323368 (M31054) nonstructural 9.7 kDa protein (put.); putative [Bovine coronavirus] (Match DDV)
>gi|465439|sp|Q04854|VSMP_CVHOC PUTATIVE SMALL MEMBRANE PROTEIN (NONSTRUCTURAL PROTEIN NS3) (NONSTRUCTURAL 9.5 KD PROTEIN). gi|476391|pir||B44275 nonstructural protein NS3 - human coronavirus (strain OC43) gi|329569 (M99576) 9.5 kDa nonstructural protein [Human coronavirus] (Match DDV)
>gi|549520|sp|P36566|YCBD_ECOLI HYPOTHETICAL 29.8 KD PROTEIN IN KDSB-KICB INTERGENIC REGION. gi|1261828 (D26440) S-adenosylmethionine-dependent methltransferase [Escherichia coli] gi|1585880|prf||2202211A Met(S-adenosyl)-dependent methyltransferase [Escherichia coli] (Match DKV)
>gi|465867|sp|P34403|YLU9_CAEEL HYPOTHETICAL 14.8 KD PROTEIN F10E9.9 IN CHROMOSOME III. (Match DTV)
>gi|119932|sp|P00229|FER1_PHYAM FERREDOXIN I. gi|65749|pir||FEFW1 ferredoxin [2Fe-2S] I - Virginian pokeweed (Match DIV)
>gi|119959|sp|P14938|FER3_RAPSA FERREDOXIN, LEAF L-A. (Match DMV)
>gi|130608|sp|P05960|POL_HV1C4 POL POLYPROTEIN (PROTEASE (RETROPEPSIN) ; REVERSE TRANSCRIPTASE ; RIBONUCLEASE H. (Match DEV)
>gi|131765|sp|P21760|QSP_CHICK QUIESCENCE-SPECIFIC PROTEIN PRECURSOR (P20K) (CH21 PROTEIN). gi|86417|pir||A30230 quiescence-specific protein precursor - chicken (Match DEV)
>gi|208939 (M14181) preproparathyroid hormone [Artificial gene] gi|209049 (M14182) preproparathyroid hormone [Artificial gene] gi|209052 (M14183) preproparathyroid hormone [Artificial gene] (Match DMV)

FIG. 8L-1

>gil209048 (M14182) synthetic preproparathyroid hormone [Artificial gene]
gil209051 (M14183) synthetic preproparathyroid hormone [Artificial gene] (Match DMV)
>gil344735 (A04054) MDV gB gene product [unidentified] gil412763 (A06147) gB gene product [unidentified] (Match DAV)
>gil221553 (D10134) NS-5 [Hepatitis C virus] (Match DPV)
>gil234099|bbs|52140 NS3 protein [bovine enteritic coronavirus BECV, strain F15, Peptide, 84 aa] (Match DDV)
>gil256415|bbs|114657 VP3=major structural polypeptide {N-terminal} [infectious flacherie virus IFV, silkworm Bombyx mori, Peptide Partial, 15 aa] (Match DIV)
>gil454753 (U04469) polymerase [Desert Shield virus] (Match DGV)
>gil1364135|pir||E49600 probable aphid transmission factor - soybean dwarf virus gil436022 (L24049) coat protein [Soybean dwarf virus] (Match DLV)
>gil471720 (U01886) gB homolog [Gallid herpesvirus 2] (Match DAV)
>gil323678 (M60583) ORF 1; putative [Densovirus of Bombyx type 1] (Match DYV)
>gil305785 (L19242) glycoprotein 120 [Human immunodeficiency virus type 1] (Match DPV)
>gil385141 (L23451) nonstructural protein 5 [Hepatitis C virus type 2b] (Match DPV)
>gil385143 (L23452) nonstructural protein 5 [Hepatitis C virus type 2b] (Match DPV)
>gil385149 (L23455) nonstructural protein 5 [Hepatitis C virus type 2b] (Match DPV)
>gil332344 (M24712) hemagglutinin-neuraminidase [Newcastle disease virus] (Match DGV)
>gil332346 (M24713) hemagglutinin-neuraminidase [Newcastle disease virus] (Match DKV)
>gil332348 (M24714) hemagglutinin-neuraminidase [Newcastle disease virus] (Match DGV)
>gil332350 (M24715) hemagglutinin-neuraminidase [Newcastle disease virus] (Match DGV)
>gil332360 (M22110) hemagglutinin-neuraminidase [Newcastle disease virus] (Match DGV)
>gil457315 (L23828) RNA polymerase [Norwalk virus] (Match DGV)
>gil295510 (L07937) 37 kDa protein [Soil-borne wheat mosaic virus] (Match DSV)

FIG. 8L-2

>gi|465058|sp|P33878|VB17_VARV PROTEIN B17. gi|419242|pir||I36856 B18L protein - variola virus (strain India-1967) gi|628217|pir||S46875 gene B17L protein (COP) - variola virus gi|439093 (L22579) homolog of vaccinia virus CDS B17L; putative [variola major virus] gi|457077 (X69198) pid:g457077 [Variola virus] gi|516436 (X67117) B17L COP gene product [Variola virus] gi|885783 (U18339) D6L [Variola virus] gi|885845 (U18341) B15L [Variola virus] gi|1150675 (X72086) ORF17L; B18L in citation [3] [Variola virus] gi|745309|prf||2015436HK B18L gene [Variola major virus] (Match DNV)
>gi|138975|sp|P15775|VSMP_CVBF PUTATIVE SMALL MEMBRANE PROTEIN (NONSTRUCTURAL PROTEIN NS3) (NONSTRUCTURAL 9.7 KD PROTEIN). gi|74875|pir||MNIHB3 nonstructural protein NS3 - bovine coronavirus (strain F15) gi|58686 (X51347) NS3 protein (AA 1-84) [Bovine coronavirus] (Match DDV)
>gi|138976|sp|P15779|VSMP_CVBM PUTATIVE SMALL MEMBRANE PROTEIN (NONSTRUCTURAL PROTEIN NS3) (NONSTRUCTURAL 9.7 KD PROTEIN). gi|418984|pir||D46346 nonstructural protein NS3 - bovine coronavirus (strain Mebus) gi|323368 (M31054) nonstructural 9.7 kDa protein (put.); putative [Bovine coronavirus] (Match DDV)
>gi|465439|sp|Q04854|VSMP_CVHOC PUTATIVE SMALL MEMBRANE PROTEIN (NONSTRUCTURAL PROTEIN NS3) (NONSTRUCTURAL 9.5 KD PROTEIN). gi|476391|pir||B44275 nonstructural protein NS3 - human coronavirus (strain OC43) gi|329569 (M99576) 9.5 kDa nonstructural protein [Human coronavirus] (Match DDV)
>gi|549520|sp|P36566|YCBD_ECOLI HYPOTHETICAL 29.8 KD PROTEIN IN KDSB-KICB INTERGENIC REGION. gi|1261828 (D26440) S-adenosylmethionine-dependent methltransferase [Escherichia coli] gi|1585880|prf||2202211A Met(S-adenosyl)-dependent methyltransferase [Escherichia coli] (Match DKV)
>gi|465867|sp|P34403|YLU9_CAEEL HYPOTHETICAL 14.8 KD PROTEIN F10E9.9 IN CHROMOSOME III. (Match DTV)
>gi|119932|sp|P00229|FER1_PHYAM FERREDOXIN I. gi|65749|pir||FEFW1 ferredoxin [2Fe-2S] I - Virginian pokeweed (Match DIV)
>gi|119959|sp|P14938|FER3_RAPSA FERREDOXIN, LEAF L-A. (Match DMV)
>gi|130608|sp|P05960|POL_HV1C4 POL POLYPROTEIN (PROTEASE (RETROPEPSIN) ; REVERSE TRANSCRIPTASE ; RIBONUCLEASE H. (Match DEV)
>gi|131765|sp|P21760|QSP_CHICK QUIESCENCE-SPECIFIC PROTEIN PRECURSOR (P20K) (CH21 PROTEIN). gi|86417|pir||A30230 quiescence-specific protein precursor - chicken (Match DEV)
>gi|208939 (M14181) preproparathyroid hormone [Artificial gene] gi|209049 (M14182) preproparathyroid hormone [Artificial gene] gi|209052 (M14183) preproparathyroid hormone [Artificial gene] (Match DMV)

FIG. 8M-1

>gil209048 (M14182) synthetic preproparathyroid hormone [Artificial gene]
gil209051 (M14183) synthetic preproparathyroid hormone [Artificial gene] (Match DMV)
>gil344735 (A04054) MDV gB gene product [unidentified] gil412763 (A06147) gB gene product [unidentified] (Match DAV)
>gil221553 (D10134) NS-5 [Hepatitis C virus] (Match DPV)
>gil234099|bbs|52140 NS3 protein [bovine enteritic coronavirus BECV, strain F15, Peptide, 84 aa] (Match DDV)
>gil256415|bbs|114657 VP3=major structural polypeptide {N-terminal} [infectious flacherie virus IFV, silkworm Bombyx mori, Peptide Partial, 15 aa] (Match DIV)
>gil454753 (U04469) polymerase [Desert Shield virus] (Match DGV)
>gil1364135|pir||E49600 probable aphid transmission factor - soybean dwarf virus gil436022 (L24049) coat protein [Soybean dwarf virus] (Match DLV)
>gil471720 (U01886) gB homolog [Gallid herpesvirus 2] (Match DAV)
>gil323678 (M60583) ORF 1; putative [Densovirus of Bombyx type 1] (Match DYV)
>gil305785 (L19242) glycoprotein 120 [Human immunodeficiency virus type 1] (Match DPV)
>gil385141 (L23451) nonstructural protein 5 [Hepatitis C virus type 2b] (Match DPV)
>gil385143 (L23452) nonstructural protein 5 [Hepatitis C virus type 2b] (Match DPV)
>gil385149 (L23455) nonstructural protein 5 [Hepatitis C virus type 2b] (Match DPV)
>gil332344 (M24712) hemagglutinin-neuraminidase [Newcastle disease virus] (Match DGV)
>gil332346 (M24713) hemagglutinin-neuraminidase [Newcastle disease virus] (Match DKV)
>gil332348 (M24714) hemagglutinin-neuraminidase [Newcastle disease virus] (Match DGV)
>gil332350 (M24715) hemagglutinin-neuraminidase [Newcastle disease virus] (Match DGV)
>gil332360 (M22110) hemagglutinin-neuraminidase [Newcastle disease virus] (Match DGV)
>gil457315 (L23828) RNA polymerase [Norwalk virus] (Match DGV)
>gil295510 (L07937) 37 kDa protein [Soil-borne wheat mosaic virus] (Match DSV)

FIG. 8M-2

>gil433113 (U03762) multigene family 360 protein [African swine fever virus] (Match DTV)
>gil76494|pir||S09759 hypothetical protein TRL10 precursor - human cytomegalovirus (strain AD169) gil59601 (X17403) HCMVTRL10 = IRL10 (AA 1-171) [Human cytomegalovirus] (Match DNV)
>gil211503 (M55644) marker protein [Gallus gallus] (Match DEV)
>gil576796 (M25784) quiescence-specific protein [Gallus gallus] (Match DEV)
>gil509165 (X70945) cellular retinoic acid binding protein I [Ambystoma mexicanum] (Match DDV)
>gil227060|prf||1613430A rimK assocd ORF [Escherichia coli] (Match DQV)
>gil76336|pir||COSJS 1G2 protein - bracket fungus (Schizophyllum commune) gil224150|prf||1011193A 1G2 gene ORF [Schizophyllum commune] (Match DPV)
>gil1346547|sp|P48040|ML1A_SHEEP MELATONIN RECEPTOR TYPE 1A (MEL-1A-R). gil602132 (U14109) Mel-1a melatonin receptor [Ovis aries] (Match DSV)
>gil625362|pir||A61338 flavodoxin - Clostridium pasteurianum (fragment) (Match DVV)
>gil625983|pir||JC2251 S-locus-specific glycoprotein S8 precursor - field mustard gil1304011 (D84468) SLG8 [Brassica campestris] (Match DLV)
>gil628958|pir||S45092 cops protein - Streptococcus pyogenes gil1333835 (X66468) copS gene product [Streptococcus pyogenes] (Match DFV)
>gil629545|pir||S40470 protein kinase 4, mitogen-activated - Arabidopsis thaliana (Match DSV)
>gil1076461|pir||S51139 S locus glycoprotein - wild cabbage gil624941 (X79431) S locus glycoprotein [Brassica oleracea] (Match DLV)
>gil601812|bbs|151834 (S72011) P14=small low-abundant nonstructural protein [bacteriophage, phi 6, Peptide, 62 aa] (Match DGV)
>gil632906|bbs|152232 RNA polymerase [human enteric calicivirus HCV, Peptide Partial, 54 aa] (Match DGV)
>gil676884 (D29681) The expression is induced by Pi starvation. [Nicotiana tabacum] gil1094819|prf||2106387C Al-induced protein [Nicotiana tabacum] (Match DRV)
>gil729540|sp|P80348|FUC2_RAT FUCTININ 2 (FUCOSYLTRANSFERASE INHIBITOR 2). gil639583|bbs|155067 fuctinin peptide 2=fucosyltransferase inhibitor {N-terminal} [rats, small intestinal mucosa, Peptide Partial, 22 aa] (Match DEV)
>gil755077 (L34630) membrane protein [Synechocystis sp.] gil1653000 (D90910) Mn transporter MntB [Synechocystis sp.] (Match DQV)
>gil765093 (D50053) ORF5 [Orgyia pseudotsugata nuclear polyhedrosis virus] gil1584397|prf||2122421B ORF 5 [Orgyia pseudotsugata nuclear polyhedrosis virus] (Match DKV)

FIG. 8N-1

>gil765256|bbs|156682 (S73813) lymphoid cell activation antigen, CD39=guanosine diphosphatase homolog [human, B lymphoblastoid cell line, MP-1, Peptide, 510 aa] (Match DMV)
>gil1083916|pir||JC2572 hypothetical 18K protein - Leuconostoc oenos phage L10 gil806612 (L13035) ORFA [Bacteriophage L10] (Match DDV)
>gil808689 (M19004) unknown protein [Saimirine herpesvirus 1] (Match DWV)
>gil261755|bbs|122153 aconitase, iron-responsive element binding protein, IRE-BP {EC 4.2.1.3} [cattle, liver cytosol, Peptide Partial, 11 aa, segment 4 of 6] (Match DVV)
>gil544869|bbs|142782 beta-glucosidase [Hordeum vulgare=barley, Sofia, Peptide Partial, 15 aa, segment 2 of 6] (Match DGV)
>gil400168|sp||LCAT_PIG_10 [Segment 10 of 11] PHOSPHATIDYLCHOLINE-STEROL ACYLTRANSFERASE PRECURSOR (LECITHIN-CHOLESTEROL ACYLTRANSFERASE) (PHOSPHOLIPID-CHOLESTEROL ACYLTRANSFERASE). (Match DPV)
>gil400776|sp||PHLD_HUMAN_6 [Segment 6 of 8] PHOSPHATIDYLINOSITOL-GLYCAN-SPECIFIC PHOSPHOLIPASE D (PI-G PLD) (GLYCOPROTEIN PHOSPHOLIPASE D). (Match DXV)
>gil860940 (X78951) core protein [Hepatitis C virus] (Match DGV)
>gil881414 (U27512) trichocyst matrix protein T4 [Paramecium tetraurelia] gil881416 (U27513) trichocyst matrix protein T4 [Paramecium tetraurelia] (Match DKV)
>gil881418 (U27514) trichocyst matrix protein T4 [Paramecium tetraurelia] (Match DKV)
>gil1361418|pir||S57659 argininosuccinate synthase (EC 6.3.4.5) - Streptomyces clavuligerus gil886906 (Z49111) argininosuccinate synthetase [Streptomyces clavuligerus] gil1586511|prf||2204224A argininosuccinate synthetase [Streptomyces clavuligerus] (Match DLV)
>gil899227 (X03170) SLSG (COOH end); pid:e188274 [Brassica oleracea] (Match DLV)
>gil913953|bbs|164394 threonine dehydrogenase, TDH {N-terminal} {EC 1.1.1.103} [Clostridium sticklandii, DSM 519T, ATCC 12662, Peptide Partial, 30 aa] (Match DNV)
>gil65750|pir||FEFWF ferredoxin [2Fe-2S] I - food pokeweed (Match DIV)

FIG. 8N-2

>gi|419454|pir||H46328 hemagglutinin-neuraminidase (EC 3.2.1.-) - Newcastle disease virus (strain AUS/32) (Match DGV)
>gi|419463|pir||I46328 hemagglutinin-neuraminidase (EC 3.2.1.-) - Newcastle disease virus (strain MIY/51) (Match DKV)
>gi|419461|pir||B36829 hemagglutinin-neuraminidase (EC 3.2.1.-) - Newcastle disease virus (strain ITA/45) (Match DGV)
>gi|81707|pir||JX0082 ferredoxin [2Fe-2S] A, leaf - radish (Match DVV)
>gi|81752|pir||S06631 lectin - coral tree (Match DAV)
>gi|541283|pir||B49850 chlorin reductase subunit BchX - Rhodobacter capsulatus (Match DDV)
>gi|99563|pir||A28965 ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - spinach (fragments) (Match DTV)
>gi|81701|pir||S04906 S-locus-specific glycoprotein S29-2 precursor - wild cabbage (fragment) (Match DLV)
>gi|89156|pir||A05311 apolipoprotein A-I - pig (fragment) (Match DRV)
>gi|89263|pir||B29544 phosphatidylcholine--sterol O-acyltransferase (EC 2.3.1.43) peptide B - pig (fragment) (Match DPV)
>gi|911252|pat|US|5411941|10 Sequence 10 from Patent US 5411941 gi|1608026|pat|US|5508263|10 Sequence 10 from patent US 5508263 (Match DMV)
>gi|911989|pat|US|5422424|1 Sequence 1 from patent US 5422424 (Match DLV)
>gi|106794|pir||S17112 interferon alpha/beta receptor - human (Match DFV)
>gi|1175245|sp|P43996|Y421_HAEIN HYPOTHETICAL PROTEIN HI0421. gi|1074400|pir||E64007 hypothetical protein HI0421 - Haemophilus influenzae (strain Rd KW20) gi|1573398 (U32725) H. influenzae predicted coding region HI0421 [Haemophilus influenzae] (Match DKV)
>gi|1176329|sp|P44812|YIIU_HAEIN HYPOTHETICAL PROTEIN HI0668. gi|1074476|pir||D64156 hypothetical protein HI0668 - Haemophilus influenzae (strain Rd KW20) gi|1573669 (U32750) hypothetical [Haemophilus influenzae] (Match DNV)
>gi|927494 (X89861) 9.6 kDa nonstructural protein [Porcine hemagglutinating encephalomyelitis coronavirus] (Match DDV)
>gi|927497 (X89863) 9.6 kDa nonstructural protein [Porcine hemagglutinating encephalomyelitis coronavirus] (Match DDV)
>gi|927500 (X89862) 9.6 kDa nonstructural protein [Porcine hemagglutinating encephalomyelitis coronavirus] (Match DDV)
>gi|947124|bbs|163644 ferredoxin component a1 [Raphanus sativus var. longipinnatus=Chinese radish, leaves, seedlings, Peptide, 96 aa] (Match DVV)
>gi|1363938|pir||S53870 metalloproteinase-3 tissue inhibitor - human gi|957310|bbs|165606 hTIMP-3=tissue inhibitor of metalloproteinase-3 {N-terminal} [human, Peptide Partial, 18 aa] (Match DIV)

FIG. 8O-1

>gil971666 (F14634) rho protein dissociation inhibitor homolog [Sus scrofa] (Match DIV)
>gil995573 (U03772) putative transposase [Acinetobacter calcoaceticus] (Match DTV)
>gil995574 (U03772) ORF2 gene product [Acinetobacter calcoaceticus] (Match DTV)
>gil998292 (U33482) ependymin [Gasteropelecus sp.] (Match DGV)
>gil998306 (U33487) ependymin [Nannobrycon sp.] (Match DGV)
>gil1346543|sp|P49285|ML1A_CHICK MELATONIN RECEPTOR TYPE 1A (MEL-1A-R). gil1000104 (U31820) Mel-1a melatonin receptor [Gallus gallus] (Match DSV)
>gil1001110 (D64001) hypothetical protein [Synechocystis sp.] (Match DSV)
>gil1001172 (D64001) hypothetical protein [Synechocystis sp.] (Match DGV)
>gil1001295 (D64006) hypothetical protein [Synechocystis sp.] (Match DPV)
>gil1016694 (U33011) urease accessory protein G [Mycobacterium tuberculosis] gil1583729|prf||2121356E urease:SUBUNIT=G [Mycobacterium tuberculosis] (Match DGV)
>gil1042011|bbs|169021 (S78693) cyclic AMP response element-binding protein-1 alpha isoform= alpha CREB-1 {alternatively spliced, internal fragment} [human, placenta, Peptide Partial, 21 aa] (Match DSV)
>gil1050760 (X83665) ribulose-1,5-bisphosphate carboxylase [Rogiera suffrutenscens] (Match DPV)
>gil1051157 (X91985) glycoprotein 100 [Marek disease virus type 1] (Match DAV)
>gil1052601 (X82442) pid:e122803 [Gallus gallus] (Match DGV)
>gil1061312 (M87661) nonstructural polyprotein [Norwalk calicivirus] (Match DGV)
>gil1351660|sp|Q09907|YAJ7_SCHPO HYPOTHETICAL 40.2 KD PROTEIN C30D11.07 IN CHROMOSOME I. gil1065894 (Z67961) unknown [Schizosaccharomyces pombe] (Match DLV)
>gil1353146|sp|Q09637|YR11_CAEEL PROBABLE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE T27D1.1 (PPIASE) (ROTAMASE). (Match DLV)

FIG. 8O-2

>gil1071799|pir||PA0003 nucleoside-diphosphate kinase (EC 2.7.4.6) - Arabidopsis thaliana (fragment) (Match DGV)
>gil1083351|pir||PC2239 heat shock protein, high-molecular-mass 105B - mouse (fragments) (Match DMV)
>gil1083905|pir||A55209 H transfer determinant A - plasmid R478 gil1326033 (L20341) IncHI2 transfer repressor [Plasmid R478] (Match DEV)
>gil1100235 (L48985) resolvase [Pseudomonas syringae] (Match DKV)
>gil1122533 (U39944) BELL1 [Arabidopsis thaliana] (Match DIV)
>gil1176915|sp|P42955|YSLB_BACSU HYPOTHETICAL 17.3 KD PROTEIN IN LYSC 3'REGION. gil1129090 (J03294) ORF; putative [Bacillus subtilis] (Match DPV)
>gil1139612 (U43400) structural phosphoprotein [Human herpesvirus 7] (Match DVV)
>gil1146150 (L43365) fiber protein [Human adenovirus type 2] (Match DGV)
>gil1150923 (X94677) major DNA binding protein [Bovine herpesvirus 1] gil1491628|gnl||PID|e258523 (Z78205) UL29 [Bovine herpesvirus 1] (Match DMV)
>gil1160339 (U21000) MerR [Pseudomonas stutzeri] gil1586135|prf||2203290A merR gene [Pseudomonas stutzeri] (Match DAV)
>gil1163120 (U43537) ORF1; putative ABC excision nuclease repair protein [Streptomyces argillaceus] (Match DAV)
>gil1164905 (X83637) ribulose-1,5-bisphosphate carboxylase, large subunit [Gardenia thunbergia] (Match DKV)
>gil1171462|bbs|171023 SnaA=pristinamycin IIA synthase 50 kda subunit {N-terminal, internal fragment} [Streptomyces pristinaespiralis, SP92, ATCC 25486, Peptide Partial, 20 aa, segment 2 of 2] (Match DFV)
>gil1173549 (U31208) NADH dehydrogenase type 1 subunit [Anabaena sp.] (Match DWV)
>gil1181520 (U42580) A357L [Paramecium bursaria Chlorella virus 1] (Match DFV)
>gil1172748|sp|P36672|PTTB_ECOLI PTS SYSTEM, TREHALOSE-SPECIFIC IIBC COMPONENT (EIIBC-TRE) (TREHALOSE-PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT) (EII-TRE). (Match DIV)
>gil1204170 (Z69729) unknown [Schizosaccharomyces pombe] (Match DNV)
>gil1213262 (Z69795) unknown [Schizosaccharomyces pombe] (Match DSV)
>gil1213627 (X95939) type I interferon receptor [Ovis aries] (Match DSV)
>gil1220217 (U49425) Lucilia cuprina beta esterase-related carboxylesterase (Lc79) gene, partial cds [Lucilia cuprina] (Match DGV)
>gil1225955|gnl||PID|e228613 (Z70177) homologous to yqbR of the skin element [Bacillus subtilis] (Match DKV)

FIG. 8P-1

>gil1235828 (U11972) emml gene product [Streptococcus pyogenes] (Match DEV)
>gil1235842 (U11998) emml gene product [Streptococcus pyogenes] (Match DTV)
>gil1236788 (L07418) polyprotein [Southampton virus] (Match DGV)
>gil1244418 (U26382) VP7 [group A rotavirus] (Match DRV)
>gil1246922|gnl|PID|e199301 (A27292) 21B4 [Babesia bovis] (Match DFV)
>gil1254543|pat|US|5486595|8 Sequence 8 from patent US 5486595 (Match DTV)
>gil1262126|gnl|PID|e235301 (Z70601) nonstructural protein 1 [Erythrovirus B19] (Match DLV)
>gil1293022 (U50250) ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit [Panax quinquefolius] (Match DVV)
>gil1304009 (D84469) SLG12 [Brassica campestris] (Match DLV)
>gil1350495 (L47606) ABA-responsive and embryogenesis-associated gene; lea-like protein [Picea glauca] (Match DYV)
>gil1360115|gnl|PID|e213272 (Z68147) glycoprotein B equivalent [Phocine herpesvirus type 1] (Match DEV)
>gil1352474|sp|P80507|IPYR_SYNY3 INORGANIC PYROPHOSPHATASE (PYROPHOSPHATE PHOSPHO-HYDROLASE) (PPASE). (Match DRV)
>gil1360894|pir||S54285 phosphoglycerate kinase - Thermotoga maritima (Match DGV)
>gil1399179 (U49426) 120 kDa immunodominant surface protein [Ehrlichia chaffeensis] (Match DIV)
>gil1399491 (U49666) Glp repressor [Pseudomonas aeruginosa] (Match DLV)
>gil1435070|gnl|PID|e253922 (X99085) integrase [Ascobolus immersus] (Match DYV)
>gil1458198 (U63197) helicase [Hepatitis GB virus C] gil1458200 (U63198) helicase [Hepatitis GB virus C] gil1458216 (U63206) helicase [Hepatitis GB virus C] gil1458218 (U63207) helicase [Hepatitis GB virus C] gil1458222 (U63209) helicase [Hepatitis GB virus C] (Match DSV)
>gil1458202 (U63199) helicase [Hepatitis GB virus C] (Match DSV)

FIG. 8P-2

>gil1458204 (U63200) helicase [Hepatitis GB virus C] (Match DSV)
>gil1458206 (U63201) helicase [Hepatitis GB virus C] (Match DSV)
>gil1458208 (U63202) helicase [Hepatitis GB virus C] gil1458220 (U63208) helicase [Hepatitis GB virus C] (Match DSV)
>gil1458210 (U63203) helicase [Hepatitis GB virus C] (Match DSV)
>gil1458212 (U63204) helicase [Hepatitis GB virus C] (Match DSV)
>gil1458214 (U63205) helicase [Hepatitis GB virus C] (Match DSV)
>gil1458224 (U63210) helicase [Hepatitis GB virus C] (Match DSV)
>gil1480344|gnl|PID|e254807 (X99405) glucose-6-phosphate dehydrogenase [Nicotiana tabacum] (Match DLV)
>gil1311403|pdb|1AUS|L Activated Unliganded Spinach Rubisco Mol_id: 1; Molecule: Ribulose Bisphosphate CarboxylaseOXYGENASE; Chain: L, S; Synonym: Rubisco; Ec: 4.1.1.39; Heterogen: Carbon Dioxide; Heterogen: Magnesium (Match DTV)
>gil1491736|gnl|PID|e223596 (X95287) archaeal ABC-transporter system [Methanosarcina mazeii] (Match DAV)
>gil1592296 (U67506) M. jannaschii predicted coding region MJ0568 [Methanococcus jannaschii] (Match DKV)
>gil1518406|gnl|PID|e220405 (Z69198) ribulose-1,5-bisphosphate carboxylase, large subunit [Triteleia bridgesii] (Match DLV)
>gil1518698 (U61753) C3-3 [Oncorhynchus mykiss] (Match DVV)
>gil1526499 (D87414) MHC class II histocompatibility antigen [Sus scrofa] (Match DTV)
>gil1526505 (D87417) MHC class II histocompatibility antigen [Sus scrofa] (Match DTV)
>gil1526525 (D87427) MHC class II histocompatibility antigen [Sus scrofa] (Match DTV)
>gil1526527 (D87428) MHC class II histocompatibility antigen [Sus scrofa] (Match DTV)
>gil1526529 (D87429) MHC class II histocompatibility antigen [Sus scrofa] (Match DTV)
>gil1526531 (D87430) MHC class II histocompatibility antigen [Sus scrofa] (Match DTV)
>gil1526533 (D87431) MHC class II histocompatibility antigen [Sus scrofa] (Match DTV)
>gil1545998 (U60650) polyprotein [Drosophila x virus] (Match DIV)
>gil1553002 (U65978) interferon alpha/beta receptor-1 [Ovis aries] (Match DSV)
>gil1567698|gnl|PID|e254689 (A32883) thrombin inhibitor protein [Rhodnius prolixus] gil1610446|pat|US|5523287|5 Sequence 5 from patent US 5523287 (Match DPV)

FIG. 8Q-1

>gi|1567700|gnl|PID|e254629 (A32885) thrombin inhibitor protein [Rhodnius prolixus] gi|1610447|pat|US|5523287|7 Sequence 7 from patent US 5523287 (Match DPV)
>gi|1567702|gnl|PID|e254631 (A32887) thrombin inhibitor protein [Rhodnius prolixus] gi|1610448|pat|US|5523287|9 Sequence 9 from patent US 5523287 (Match DPV)
>gi|1567704|gnl|PID|e254632 (A32889) thrombin inhibitor protein [Rhodnius prolixus] gi|1610449|pat|US|5523287|11 Sequence 11 from patent US 5523287 (Match DPV)
>gi|1567706|gnl|PID|e254633 (A32891) thrombin inhibitor protein [Rhodnius prolixus] gi|1610450|pat|US|5523287|13 Sequence 13 from patent US 5523287 (Match DPV)
>gi|1567708|gnl|PID|e254634 (A32893) thrombin inhibitor protein [Rhodnius prolixus] gi|1610451|pat|US|5523287|15 Sequence 15 from patent US 5523287 (Match DPV)
>gi|1567710|gnl|PID|e254691 (A32895) thrombin inhibitor protein [Rhodnius prolixus] gi|1610452|pat|US|5523287|17 Sequence 17 from patent US 5523287 (Match DPV)
>gi|1575524 (U65005) structural phosphoprotein [Human herpesvirus 7] (Match DVV)
>gi|1607344|pat|US|5500347|2 Sequence 2 from patent US 5500347 (Match DKV)
>gi|1607345|pat|US|5500347|3 Sequence 3 from patent US 5500347 (Match DKV)
>gi|1607346|pat|US|5500347|4 Sequence 4 from patent US 5500347 (Match DKV)
>gi|1607348|pat|US|5500347|6 Sequence 6 from patent US 5500347 (Match DKV)
>gi|1607349|pat|US|5500347|7 Sequence 7 from patent US 5500347 (Match DKV)
>gi|1608953|pat|US|5510461|9 Sequence 9 from patent US 5510461 (Match DHV)
>gi|1610343|pat|US|5521071|2 Sequence 2 from patent US 5521071 (Match DLV)
>gi|1610926|pat|US|5527773|3 Sequence 3 from patent US 5527773 (Match DKV)
>gi|1610980|pat|US|5527896|56 Sequence 56 from patent US 5527896 (Match DMV)
>gi|1610981|pat|US|5527896|57 Sequence 57 from patent US 5527896 (Match DIV)
>gi|1611666|pat|US|5539092|98 Sequence 98 from patent US 5539092 (Match DKV)
>gi|1613384|pat|US|5559008|67 Sequence 67 from patent US 5559008 (Match DDV)
>gi|1613387|pat|US|5559008|70 Sequence 70 from patent US 5559008 (Match DDV)
>gi|1587874|prf|l2207325A Ant1 gene [Aspergillus niger] (Match DEV)

FIG. 8Q-2

>gil1648974 (U66481) P1 structural protein [Hepatitis A virus] gil1648990 (U66489) P1 structural protein [Hepatitis A virus] (Match DPV)
>gil1648976 (U66482) P1 structural protein [Hepatitis A virus] gil1648986 (U66487) P1 structural protein [Hepatitis A virus] (Match DPV)
>gil1648978 (U66483) P1 structural protein [Hepatitis A virus] (Match DPV)
>gil1648980 (U66484) P1 structural protein [Hepatitis A virus] (Match DPV)
>gil1648982 (U66485) P1 structural protein [Hepatitis A virus] (Match DPV)
>gil1648984 (U66486) P1 structural protein [Hepatitis A virus] (Match DPV)
>gil1648988 (U66488) P1 structural protein [Hepatitis A virus] (Match DPV)
>gil1651445 (D90730) Hypothetical 29.8 KD protein in kdsB-kicB intergenic region [Escherichia coli] (Match DKV)
>gil1651926 (D90901) hypothetical protein [Synechocystis sp.] (Match DLV)
>gil1651969 (D90901) hypothetical protein [Synechocystis sp.] (Match DDV)
>gil1652043 (D90902) hypothetical protein [Synechocystis sp.] (Match DLV)
>gil1653351 (D90913) HlyB family [Synechocystis sp.] (Match DDV)
>gil1654110 (U14110) melatonin receptor Mel-1a [Phodopus sungorus] (Match DSV)
>gil1655822 (U59320) heat shock protein 60 [Leishmania major] (Match DEV)
>gil1657485 (U73857) similar to E. coli o765 [Escherichia coli] (Match DVV)
>gil1658269 (U74670) 120 kDa immunodominant surface protein [Ehrlichia chaffeensis] (Match DIV)

FIG. 8R

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/21861

A. CLASSIFICATION OF SUBJECT MATTER
IPC(6) : C12Q 1/68, G01N 33/53
US CL : 435/6, 435/7.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/6, 435/7.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | GALLOP et al. Applications of combinatorial technologies to drug discovery. 1. Background and peptide combinatorial libraries. Journal of Medicinal Chemistry. 29 April 1994, Vol.37 No.9, pages 1233-1251, especially pages 1237-1238 under the heading Peptides-on-Plasmids. | 3-4, 9-15 23-31 |
| X, P | Schepens et al. The neuronal nitric oxide synthase PDZ motif binds to -G(D,E)XV* carboxyterminal sequences. FEBS Letters. June 1997, Vol. 409, pages 53-56, especially Tables 2 and 3. | 1, 2, 6-8, 16-19, 21, 22 |

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

| | | | |
|---|---|---|---|
| * | Special categories of cited documents: | "T" | later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention |
| "A" | document defining the general state of the art which is not considered to be of particular relevance | "X" | document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone |
| "E" | earlier document published on or after the international filing date | "Y" | document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| "L" | document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | "&" | document member of the same patent family |
| "O" | document referring to an oral disclosure, use, exhibition or other means | | |
| "P" | document published prior to the international filing date but later than the priority date claimed | | |

| | |
|---|---|
| Date of the actual completion of the international search<br>04 FEBRUARY 1998 | Date of mailing of the international search report<br>02 MAR 1998 |
| Name and mailing address of the ISA/US<br>Commissioner of Patents and Trademarks<br>Box PCT<br>Washington, D.C. 20231<br>Facsimile No. (703) 305-3230 | Authorized officer<br>JOSEPH W. RICIGLIANO<br>Telephone No. (703) 308-0196 |

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/21861

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| A | KIM at al. Clustering of Shaker-type K+ channels by interaction with a family of membrane-associated guanylate kinases. Nature. 02 November 1995, Vol.378, pages 85-88. | 1, 2, 6-8, 17-19, 21-31 |
| X | WO 95/28426 A2 (LA JOLLA CANCER RESEARCH FOUNDATION) 26 October 95, See entire document, especially pages 9-10 references to SEQ ID NO: 1 and page 25, line 4 to page 30 line 19. | 1, 3, 5-8, 20 |

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/21861

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS (JPOABS, EPOABS, USPAT), Medline 1983 - September 1997, CAS ONLINE
DXV^/sqsp and X.+DXV^/sqsp in CAS registy; nitric oxide synthase and peptide and PDZ, DXV, D-X-V